SINAPPSI
CONNESSIONI TRA RICERCA E POLITICHE PUBBLICHE
Rivista quadrimestrale dell'Istituto nazionale per l'analisi delle politiche pubbliche

Intelligenza artificiale tra efficienza e umanità

Implicazioni dell'IA per le scienze, le tecnologie e la società

Aspetti legali, etici e di regolamentazione

Relazione tra mercato del lavoro e competenze umane

A cura di Marcella Milana e Maria Enrica Virgillito

Contributi

Cosimo Accoto
Giovanni Armillotta
Andrea Battistoni
Alberto Bortolami
Irene Brunetti
Francesco Capone
Giusy Caravello
Emilio Colombo
Laura Evangelista
Enrico Ferraris
Valentina Ferri
Antonio Finazzi Agrò
Concetta Fonzo
Filippo Giordano
Martina Gnudi
Giorgio Gosetti
Francesca Lettieri
Saverio Lovergine
Federico Maggiora
Annaleda Mazzucato
Andrei Mihai Pop
Fortunato Musella
Sara Pane
Sara Pautasso
Marco Peruzzi
Luciano Pietronero
Andrea Ricci
Gaetano Riccio
Federica Romano
Alessia Romito
Luigi Rullo
Boris Sofronic
Francesco Trentini
Marco Trombetti
Eleonora Zecca

Anno XV n. 2/2025
INAPP Editore

SINAPPSI

CONNESSIONI TRA RICERCA E POLITICHE PUBBLICHE

Rivista scientifica quadrimestrale di analisi dei fenomeni sociali e valutazione delle politiche che riguardano l'occupazione e lo sviluppo delle risorse umane, la formazione, l'inclusione e la protezione sociale.

INAPP Editore
00198 Roma - Corso d'Italia, 33
Tel. +39 06 854471
www.inapp.gov.it

Iscrizione al Tribunale di Roma
n. 420/2010 del 21/10/2010

ISBN: 978-88-543-0363-8
ISSN: 2532-8549
E-ISSN: 2611-6332

Fascicolo chiuso e pubblicato ad agosto 2025

Finito di stampare
nel mese di ottobre 2025
da Rubbettino Print
88049 Soveria Mannelli (CZ))

L'Istituto nazionale per l'analisi delle politiche pubbliche (INAPP) è un ente pubblico di ricerca che si occupa di analisi, monitoraggio e valutazione delle politiche del lavoro, delle politiche dell'istruzione e della formazione, delle politiche sociali e, in generale, di tutte le politiche economiche che hanno effetti sul mercato del lavoro. Nato il 1° dicembre 2016 a seguito della trasformazione dell'Isfol è vigilato dal Ministero del Lavoro e delle politiche sociali, l'Ente ha un ruolo strategico – stabilito dal decreto legislativo 14 settembre 2015, n. 150 – nel nuovo sistema di governance delle politiche sociali e del lavoro del Paese. L'Inapp fa parte del Sistema statistico nazionale (SISTAN) e collabora con le istituzioni europee. È Organismo Intermedio del Programma nazionale Giovani, donne e lavoro 2021-2027 del FSE+, delegato dall'Autorità di Gestione all'attuazione di specifiche azioni, ed è Agenzia nazionale del programma comunitario Erasmus+ per l'ambito istruzione e formazione professionale. È l'ente nazionale all'interno del consorzio europeo ERIC-ESS che conduce l'indagine European Social Survey. L'attività dell'Inapp si rivolge a una vasta comunità di stakeholder: ricercatori, accademici, mondo della pratica e policymaker, organizzazioni della società civile, giornalisti, utilizzatori di dati, cittadinanza in generale.

Presidente
Natale Forlani

Direttore generale
Loriano Bigi

Sommario

Intelligenza artificiale tra efficienza e umanità

Introduzione

Marcella Milana - Università degli Studi di Verona
Maria Enrica Virgillito - Scuola Superiore Sant'Anna

A settant'anni dall'introduzione del termine Intelligenza artificiale (IA) (McCarthy *et al.* 1955), l'IA ha profondamente trasformato i processi produttivi e l'erogazione dei servizi, penetrando la vita quotidiana. Ciò è stato favorito dai recenti progressi nell'Elaborazione del Linguaggio Naturale (*Natural Language Processing*), dalla raggiunta capacità computazionale e dalla possibilità di archiviazione di grandi dati. La combinazione di tali tecniche integrate permette la generazione di contenuti linguistici e visivi, da immagini, video, a contenuti musicali e testi. La recente applicazione dell'IA di 'massa', soprattutto grazie a strumenti come ChatGPT e più avanzati modelli integrativi, sta influendo i processi di elaborazione, lettura e interpretazione dei contenuti linguistici, e quindi cognitivi, a tutti i livelli dell'organizzazione sociale. Se l'apparente beneficio è quello di snellire i processi decisionali, il rischio è che la capacità dell'*intelligere*, sia completamente demandata ad 'altro da sé', un corpo macchinico estraneo che propone, suggerisce e decide sui processi umani. Da ciò, la *special issue* di questo numero della Rivista su *Intelligenza artificiale: impiego nei processi produttivi e di erogazione e accesso ai servizi, implicazioni etiche e competenze necessarie* raccoglie contributi che studiano l'applicazione di IA su varie dimensioni economico-sociali.

Le tecnologie generative dell'IA trovano applicazione in numerosi settori pubblici e privati, supportando o sostituendo lavoratrici e lavoratori in attività che vanno dalla gestione amministrativa alla produzione, fino alla ricerca scientifica e all'istruzione. Esse facilitano e ottimizzano processi e relazioni con utenti e clienti, ma influenzano anche la quantità e qualità del lavoro disponibile (Butera e De Michelis 2024), sollevando questioni di tutela contro discriminazioni, specie nei processi di selezione (Peruzzi 2024). Inoltre, recenti studi mostrano un grande interesse da parte delle organizzazioni produttive a implementare procedure di management algoritmico della forza lavoro, attraverso piattaforme supportate dall'IA, volte a efficientare la produttività e controllare i processi di lavoro, inclusa l'attribuzione di turni e di mansioni sulla base delle performance individuali (Staccioli e Virgillito 2025). Tali tecnologie pongono pertanto documentati rischi di salute e sicurezza sul lavoro. L'IA assiste inoltre la ricerca scientifica (Rice *et al.* 2024) ma può comprometterne originalità e integrità (Rahimi e Abadi 2023).

L'impiego dell'IA presenta quindi rischi e dilemmi etici. Nella Pubblica amministrazione, ad esempio, è necessario bilanciare la protezione dei diritti di cittadini e cittadine con l'efficienza (Grenci 2024). Inoltre, la relazione umano-IA può diventare problematica: se l'essere umano crea gli algoritmi, può anche esserne condizionato, come accade quando modelli generativi amplificano la disinformazione (Martire 2024; Pollicino e Dunn 2024).

Per fronteggiare queste sfide, l'Unione europea ha adottato il Regolamento (UE) 2024/1689 che definisce obblighi etici per sviluppatori e operatori, all'interno di un quadro più ampio di misure per garantire sicurezza e diritti fondamentali. Inoltre, singoli Paesi, europei e non (ad esempio Germania, Stati Uniti e Cina), hanno predisposto strategie nazionali per integrare l'IA nei processi di istruzione e formazione rivolti a giovani e adulti al fine di prepararli all'uso dell'AI (Milana *et al.* 2024).

I contributi che compongono questo volume dialogano in diversa misura con le questioni poc'anzi declinate, ruotando attorno a tre assi tematici principali.

Un primo asse tematico si concentra sulle **implicazioni dell'IA per le scienze, le tecnologie e la società**.

Nel saggio di Cosimo Accoto si presentano tre approcci interpretativi dell'IA, radicati in diversi ambiti scientifico-disciplinari. L'approccio 'strumentale', che vede l'IA come artefatto tecnico, è il più diffuso. Tuttavia, meritano attenzione anche l'approccio 'strutturale', che considera l'IA come nuova architettura produttiva che ridefinisce le interazioni socioeconomiche, e quello 'istituzionale', secondo cui l'IA crea nuove forme di istituzionalità socio-politiche, trasformando istituzioni classiche come mercati, burocrazie e Stati nazionali.

Di contro, Luciano Pietronero e Marco Trombetti concentrano l'attenzione sulle possibilità strumentali non ancora del tutto esplorate nel contesto italiano. Partendo dalla propria esperienza, discutono due questioni apparentemente distinte: la necessità di grandi reti neurali e dati massicci per avanzare nella

traduzione automatica; e la necessità di algoritmi capaci di selezionare pochi dati per garantire accuratezza nelle analisi economiche. Da tale discettazione nasce la consapevolezza della necessità di creare e sostenere, in Italia, una cultura interdisciplinare che possa permettere un approccio originale allo sviluppo di sistemi complessi per accrescere le potenzialità dell'IA.

Sul tema della ricerca scientifica, consapevoli della non neutralità degli strumenti di IA generativa, Filippo Giordano e Saverio Lovergine considerano come organismi internazionali, ma soprattutto le università, stiano agendo al fine di garantirne un uso che non infici l'integrità della ricerca prodotta. Restringendo l'attenzione a 20 delle migliori università europee (secondo più ranking disponibili), ed esaminando le linee guida in tema di IA e ricerca scientifica della metà di queste, l'analisi mostra come le università incoraggino l'uso dell'IA generativa entro limiti di trasparenza ed etica.

Sempre sull'IA generativa, Alberto Bortolami e colleghi indagano le sfide nella progettazione sociale, vista come processo decisionale collettivo e politico. Le sfide etiche e politiche derivano dal fatto che l'IA ridefinisce molte funzioni del progettista sociale (dall'analisi dei fabbisogni alla valutazione d'impatto). È dunque necessario un approccio che integri IA generativa e creatività umana, preservando autonomia e responsabilità del progettista verso le comunità.

Similmente, Sara Pane esamina le trasformazioni nella comunicazione pubblica europea dovute all'uso dell'IA generativa. Dai casi GPT@EC ed Eurodesk Chatbot emergono opportunità come l'ottimizzazione dei processi e la personalizzazione dell'interazione con i cittadini, ma anche rischi legati a infrastrutture digitali insufficienti, competenze pubbliche limitate e *bias* algoritmici. Si rende pertanto necessario garantire un approccio umano-centrico, di cui trasparenza, inclusività e responsabilità sono aspetti fondanti e imprescindibili.

Il secondo asse tematico riguarda gli **aspetti legali, etici e di regolamentazione dell'IA**.

Laura Evangelista, Concetta Fonzo ed Eleonora Zecca analizzano le possibili violazioni dei diritti fondamentali (informazione, non-discriminazione, privacy) tramite l'IA. Ne segue una disamina di come e in che misura la legislazione esistente in Europa e negli Stati Uniti affronta, limita e/o punisce tali violazioni. Nel confronto emerge come la legislazione europea adotti un approccio basato sul rischio, bilanciando innovazione e tutela, con obblighi severi per sistemi ad alto rischio; mentre gli Stati Uniti adottano un approccio frammentato e decentralizzato, influenzato da interessi di parte. L'auspicio è pertanto lo sviluppo futuro di un ecosistema normativo internazionale per garantire equità, affidabilità e sicurezza a livello globale.

Andrei Mihai Pop ed Enrico Ferraris approfondiscono l'approccio umano-centrico europeo, individuandolo come base per un 'nuovo' diritto fondamentale allo *human first*. Nel contesto della Pubblica amministrazione, dove il rapporto di dipendenza tra cittadini e Stato è marcato, gli Autori richiamano procedimenti amministrativi italiani che hanno analizzato l'uso degli algoritmi, evidenziando come la giurisprudenza ha cercato e continua a cercare di orientare la transizione della Pubblica amministrazione nel contesto della rivoluzione tecnologica dell'Intelligenza artificiale.

Sara Pautasso sposta l'attenzione al contesto industriale, dove la robotica collaborativa si sta imponendo nell'automazione del lavoro. Preoccupata per le implicazioni etiche, di sicurezza e relazioni socio-professionali derivanti dall'impiego dei 'cobot'(robot collaborativi progettati per operare in prossimità e sinergia con gli esseri umani), l'Autrice analizza tre dimensioni critiche dell'interazione uomo-macchina: il controllo temporale come potere, la tensione tra collaborazione e competizione, e l'autonomia e dignità umana. Ciò richiama l'importanza di nuove competenze per l'uso dei cobot e la prevenzione di nuove forme di subordinazione.

Tornando alla Pubblica amministrazione, Fortunato Musella e Luigi Rullo osservano la difficoltà del settore pubblico italiano nel tenere il passo del privato nelle politiche digitali. Basandosi sul concetto di *policy capacity* (mutuato dalla letteratura internazionale per qui intendere il set di competenze, abilità e risorse delle amministrazioni per progettare e perseguire obiettivi di politica pubblica), gli Autori leggono criticamente l'impatto che l'IA ha sulla Pubblica amministrazione. L'invito è a considerare il 'di più' che l'interazione uomo-macchina può offrire per innovare l'amministrazione pubblica, evidenziando le carenze e come colmarle per potenziarne le capacità di policy.

Giorgio Gosetti e Marco Peruzzi adottano un'analisi multidisciplinare sull'IA e i processi produttivi, con attenzione ai processi di innovazione nel lavoro. Gli Autori, con un approccio di sociologia del lavoro, evidenziano come l'introduzione dell'AI stia trasformando processi, contenuti lavorativi e qualità della vita, mentre da un punto di vista giuslavoristico, propongono un'architettura regolativa per governare tali cambiamenti, che includa l'identificazione e la regolamentazione delle tutele (individuali e collettive) necessarie a lavoratori e lavoratrici.

A chiusura di questo asse tematico, Gaetano Riccio torna sulla Pubblica amministrazione dal punto di vista normativo. Pur riconoscendo le potenzialità e le sfide etiche e normative dell'IA, il contributo propone un quadro integrato per l'adozione dell'IA generativa nel settore pubblico. Questo si basa sull'armonizzazione della normativa europea esistente (*Artificial Intelligence Act* 2024 e *General Data Protection Regulation* 2018), con l'obiettivo di bilanciare innovazione e conformità normativa.

Infine, il terzo asse tematico si concentra sulla **relazione tra mercato del lavoro e competenze umane rispetto all'IA**.

Nel saggio che apre questo asse tematico, Andrea Battistoni e Valentina Ferri analizzano l'impatto dell'IA nel settore bancario, finanziario e assicurativo italiano tramite uno studio quantitativo. Gli Autori descrivono l'evoluzione dell'occupazione dal 2018 al 2023 e, con analisi econometriche, indagano la relazione tra variazioni nelle attivazioni contrattuali e l'esposizione dei lavoratori all'IA. I risultati evidenziano un effetto negativo significativo sulle attivazioni contrattuali nelle professioni più esposte all'IA.

Sempre attraverso analisi econometriche, Irene Brunetti e Andrea Ricci esaminano la relazione tra investimenti in IA e domanda di lavoro nelle imprese italiane, utilizzando dati dell'Inapp dalla Rilevazione Imprese e lavoro. I risultati suggeriscono che l'adozione di sistemi di IA non influisce significativamente sulle nuove assunzioni (domanda effettiva), ma è associata a un lieve aumento dei posti di lavoro vacanti (domanda potenziale).

Alessia Romito e Boris Sofronic studiano i modelli linguistici di grandi dimensioni (*Large Language Models,* LLM) per rilevare il mismatch delle competenze in Italia in due sperimentazioni (2023, 2025). La prima ha coinvolto interazioni con ChatGPT per valutare il suo aggiornamento su nuove competenze e opportunità formative. La seconda, due anni dopo, ha impiegato strumenti che combinano LLM con sistemi di verifica basati su banche dati certificate, permettendo di definire principi per un uso efficace nella rilevazione del mismatch.

Conclude l'asse tematico Giovanni Armillotta e colleghi che propongono un'analisi prospettica sull'impatto futuro dell'IA sul mercato del lavoro italiano, considerando i mismatch di competenze per singola professione e confrontandoli con altri Paesi europei. Gli Autori evidenziano una correlazione positiva che indica un ruolo potenziale dell'IA nel favorire la crescita occupazionale e nel ridurre il mismatch tra competenze richieste e disponibili.

Complessivamente, il numero offre una prospettiva multidisciplinare che è volta a scandagliare le attuali ramificazioni dell'IA. Se da un punto di vista analitico si sottolinea la grande accelerazione nell'uso, meno evidenti sono i benefici diretti: di fatto le valutazioni e analisi ad oggi ancora poco si soffermano sul problema degli errori e dei *bias* cognitivi indotti dall'IA, come quelli di genere e razza (Unesco e Ircai 2024). Inoltre, poco è ancora chiaro quanto l'IA disabitui il pensiero e il ragionamento, e pertanto quanto apparenti benefici di breve periodo comportino invece danni di lungo periodo, come il disapprendere.

I rischi legati alla profilazione individuale e alla gestione del personale in termini discriminatori richiedono dei quadri normativi che vadano oltre il problema dalla data-privacy, e che diventino invece più sostantivi sotto il profilo dei diritti all'informazione (sull'effettivo uso dell'IA) e delle tutele di chi è subordinato a un algoritmo. Da ultimo, molto è ancora da studiare sul piano dell'impatto ambientale dei modelli di training, dell'uso e dell'estrazione di risorse e materiali critici per la creazione di ultrapotenti microprocessori, della deprivazione di territori dalla gestione delle proprie risorse naturali. Il problema dell'impatto ambientale dell'IA (Dhar 2020) e degli appetiti coloniali sottostanti alla gestione delle risorse naturali non possono come sempre essere relegati a questioni di eventuale successiva giustizia, ma richiedono che l'agenda politica le definisca da subito come prioritarie.

Riferimenti bibliografici

Butera F., De Michelis G. (2024), *Intelligenza artificiale e lavoro, una rivoluzione governabile*, Venezia, Marsilio

Dhar P. (2020), The carbon impact of artificial intelligence, *Nature Machine Intelligence*, n.2, pp.423-425

Grenci S.B. (2024), Artificial Intelligence Applications to Support the Automation of the Administrative. Procedure. Le Applicazioni Di Intelligenza Artificiale a supporto dell'automazione del Procedimento amministrativo, *Rivista italiana di informatica e diritto: periodico internazionale del CNR-IGSG*, 6, n.1 <DOI: https://doi.org/10.32091/RIID0139>

Martire D. (2024), Human in the loop. L'essere umano come fattore condizionante della – o condizionato dalla – intelligenza artificiale, *Rivista italiana di informatica e diritto*, 6, n.2, pp.467-481 <DOI:https://doi.org/10.32091/RIID0161>

McCarthy J., Minsky M., Rochester N., Shannon C. (1955), A proposal for Dartmouth summer research project on artificial intelligence, *AI Magazine*, 27, n.4, pp.12-14

Milana M., Brandi U., Hodge S., Hoggan-Kloubert T. (2024), Artificial intelligence (AI), conversational agents, and generative AI: implications for adult education practice and research, *International Journal of Lifelong Education*, 43, n.1, pp.1-7

Peruzzi M. (2024), La discriminazione algoritmica, *Equal. Rivista di Diritto Antidiscriminatorio*, n.1, pp.7-23

Pollicino O., Dunn P. (2024), Disinformazione e intelligenza artificiale nell'anno delle global elections: rischi (ed opportunità), *Federalismi.it. Rivista di Diritto Pubblico Italiano, comparato, europeo*, n.12, pp.iv-xxiii

Rice S., Crouse S.R., Winter S. R., Rice C. (2024), The advantages and limitations of using ChatGPT to enhance technological research, *Technology in Society*, 76, 102426

Rahimi F., Talebi Bezmin Abad A. (2023), Passive contribution of ChatGPT to scientific papers, *Annals of Biomedical Engineering*, 51, n.11, pp.2340-2350

Staccioli J., Virgillito M.E. (2025), Will your boss be an algorithm? A patent-based analysis of artificial intelligence worker management technologies and labour exposure, *Eurasian Business Review*, pp.1-31,<DOI:https://doi.org/10.1007/s40821-025-00317-7>

Unesco, Ircai (2024), *Challenging systematic prejudices: an Investigation into Gender Bias in Large Language Models*, Paris, Unesco, <https://unesdoc.unesco.org/ark:/48223/pf0000388971>

L'Intelligenza artificiale: soluzione, sistema o istituzione?

Cosimo Accoto
Università di Modena e Reggio Emilia
Massachusetts Institute of Technology Boston

In questo contributo si analizza la questione complessa dell'Intelligenza artificiale attraverso l'articolazione di tre dimensioni interpretative fondamentali: la soluzione, il sistema, l'istituzione. L'obiettivo è ampliare la comprensione di questo fenomeno oltre le attuali interpretazioni eminentemente 'strumentali' esplorando le dimensioni più 'strutturali' e, da ultimo, anche proprio 'istituzionali' delle intelligenze artificiali in sviluppo (ricognitive, predittive, generative, agentive). Dopo un excursus rapido sull'evoluzione delle storie dell'Intelligenza artificiale, esamineremo strategicamente l'approccio tecnologico, quello sistemico e quello istituente, con le rispettive caratterizzazioni e limitazioni.

This article examines the complex phenomenon of Artificial intelligence through three fundamental interpretative dimensions: the solution, the system, and the institution. The goal is to broaden the understanding of AI beyond current predominantly 'instrumental'interpretations by exploring the more 'structural'and, ultimately, 'institutional'dimensions of emerging forms of artificial intelligence (discriminative, predictive, generative, and agentive). Following a brief excursus of the histories of artificial intelligence, the article investigates the technological, structural, and institutional approaches, outlining their respective features and limitations.

DOI: 10.53223/Sinappsi_2025-02-1

Citazione
Accoto C. (2025), L'Intelligenza artificiale: soluzione, sistema o istituzione?, *Sinappsi*, XV, n.2, pp.8-17

Parole chiave
Cambiamento sociale
Intelligenza artificiale
Sistemi complessi

Keywords
Social change
Artificial intelligence
Complex systems

Introduzione

Erroneamente e a lungo circoscritto (reificato o personificato a seconda dei casi) alla dimensione tecnologica con le sue ramificazioni economiche, legali, sociali ed etiche, l'assemblaggio socio-tecnico che chiamiamo 'Intelligenza artificiale' si sta finalmente rivelando per quello che è sempre stato: una provocazione politica planetaria. Di fatto, evoca e incarna il crescere di una computazione che scala a livello terrestre (ed extraterrestre anche) diventando vettore generativo, spesso invisibile e talvolta inviso, di nuove sovranità volumetriche, di nuove visualità operazionali, di nuove scritture sintetiche, di nuove agentività macchiniche. Con opportunità da cogliere, naturalmente, ma anche con reali criticità da fronteggiare e solide immunità da costruire. Tra biosfere e tecnosfere, ora mature ora immature, saremmo allora alle prese speculativamente con la questione di un'emergente e sorprendente 'intelligenza planetaria'. Non artificiale, ma planetaria

proprio: la nuova forma d'esistenza, d'intelligenza e d'esperienza del nostro pianeta Terra. In questo senso, la domanda esistenziale, a detta degli astrobiologi, diviene allora questa: l'Intelligenza artificiale sta accadendo 'su' un pianeta o sta accadendo 'ad' un pianeta? (Frank *et al*. 2022). L'Intelligenza artificiale, dunque, è (sarà) un nuovo modo d'essere (abitato) del nostro Pianeta. Prefigura (e configura) l'ennesima ultima *terraformazione* del nostro mondo (Bratton 2019), la sua imminente e altra condizione di immanenza, di esperienza e di intelligenza. Nella stagione di una nuova ragione automatica (Bates 2024) e come per altri passaggi nella storia della civilizzazione umana si scardineranno ordini del discorso (regimi di verità/falsità documentali), modi della produzione (regimi di possibilità/impossibilità imprenditoriali), etiche della conduzione (regimi di responsabilità/irresponsabilità morali). Se questo è vero, allora è necessario che la nostra analisi di questo fenomeno complesso vada oltre e superi le attuali interpretazioni eminentemente 'strumentali' per avviarsi ad approfondire le dimensioni più 'strutturali' e, da ultimo, anche proprio 'istituzionali' dell'Intelligenza artificiale. O meglio delle molteplici intelligenze artificiali in sviluppo: ricognitive, predittive, generative, agentive e così via. Oltre alle dimensioni 'instrumentanti', dobbiamo dunque cogliere quelle 'infrastrutturanti' e poi 'istituzionalizzanti'. Questo slittamento paradigmatico dall'interpretazione dell'Intelligenza artificiale come 'soluzione' (tecnico-operativa) alla sua interpretazione come 'struttura' (economico-produttiva) e, infine, come 'istituzione' (socio-politica) rappresenta la vera sfida che oggi politica, economia, cultura e società fronteggiano, consapevolmente o meno. Questo contributo ha lo scopo di portare in evidenza questa stratificazione di ermeneutiche e significati per aiutare la nostra contemporaneità a orientare al meglio le analisi e gli interventi strategici e di governo di quest'emergenza. Emergenza nel senso insieme di criticità e novità per la civilizzazione umana. Dopo un iniziale, rapido excursus nella storia delle storie dell'Intelligenza artificiale e dopo un richiamo alla definizione giuridica in uso (e in divenire) dell'oggetto di studio, esploreremo sia pur in breve queste tre dimensioni interpretative del fenomeno.

1. Storie dell'Intelligenza artificiale

Le narrazioni storiografiche che riguardano l'Intelligenza artificiale hanno avuto in passato una connotazione soprattutto tecnico-ingegneristica. Hanno privilegiato, cioè, una focalizzazione centrata sul racconto e sulla descrizione delle evoluzioni nei paradigmi e nelle tecnicalità che, di volta in volta, centri di ricerca, laboratori di sviluppo e imprese innovative (oltre che i protagonisti umani delle storie) hanno immaginato e progettato. Anche le varie scelte di periodizzazione temporale sono indicative di questa interpretazione strumentale di quanto classicamente, da circa settant'anni, viene etichettato come 'Intelligenza artificiale'. Un naming *flashy* e *catchy* scelto al tempo (1955) per impressionare i finanziatori del seminario fondativo e per distinguersi da altre diciture già storicamente disciplinate e presidiata (es. la cibernetica). L'anno prescelto per segnare l'avvio di questa 'disciplina' (con l'introduzione dell'espressione *Artificial intelligence*) è, infatti, il 1956, anno del seminario estivo a Dartmouth durante il quale un gruppo di ricercatori e scienziati nominarono e istituirono questo dominio di ricerca. Le storiografie da quell'anno fondativo procedono poi variamente a individuare anticipatori e precursori nel tempo (e successori e continuatori) secondo un classico paradigma storiografico evolutivo lineare. Emblematica di questa prospettiva è la monumentale narrazione sviluppata nel saggio *The Quest for Artificial Intelligence* (Nilsson 2010) e impegnata a raccontare, con grande dettaglio, idee, tecniche e conquiste succedutesi dalla metà degli anni Cinquanta. Naturalmente sappiamo bene che ogni operazione storica di periodizzazione di un fenomeno rappresenta una scelta ideale e idealizzata (e ideologica sempre anche in certa misura) degli eventi e delle tracce storiografiche selezionate per individuare gli stessi. Queste ricostruzioni storiche hanno tuttavia, nel tempo, contribuito a rafforzare e spingere una (limitante) prospettiva interpretativa 'strumentale' del fenomeno dell'Intelligenza artificiale. Interpretazione che si è poi riflessa sulla più complessiva (in)capacità della società e dei suoi saperi di comprendere e orientare queste nuove ingegnerie della cognizione sintetica (Accoto 2024). Solo più recentemente lo sguardo storiografico si è opportunamente aperto a considerare l'automazione cognitiva come un più complesso fenomeno sociotecnico di lunga durata e di ampia portata. Così sono venuti emergendo orizzonti più problematizzanti e interdisciplinari. A titolo illustrativo, possiamo ricordare qui: l'orientamento ricostruttivo socio-politico (Pasquinelli 2023) oppure quello

geo-culturale (Cave e Dihal 2023) o ancora quello storico-concettuale (Donnelly 2024). Nel primo caso, quello 'socio-politico', si porta in evidenza come la cosiddetta Intelligenza artificiale riguardi più l'automazione dell'intelligenza del lavoro e delle relazioni sociali che l'imitazione dell'intelligenza biologica. Come scrive Pasquinelli (2023):

> "*What is AI? A dominant view describes it as the quest 'to solve intelligence' – a solution supposedly to be found in the secret logic of the mind or in the deep physiology of the brain, such as in its complex neural networks. In this book I argue, to the contrary, that the inner code of AI is constituted not by the imitation of biological intelligence but by the intelligence of labour and social relations. Today, it should be evident that AI is a project to capture the knowledge expressed through individual and collective behaviours and encode it into algorithmic models to automate the most diverse tasks: from image recognition and object manipulation to language translation and decision-making*" (p. 2).

Nel secondo caso, quello 'geo-culturale', l'attenzione interpretativa viene spostata ulteriormente a esplorare come il fenomeno dell'Intelligenza artificiale e delle sue ingegnerie non sia planetariamente omogeneo, ma venga declinato variamente a seconda delle diverse aree geografiche e culturali. Nelle parole di Cave e Dihal (2023):

> "*AI is now a global phenomenon. The term originated in the US, and much of the technology continues to be developed there. But those systems are being taken up around the world, and other countries are scrambling to develop their own AI industries. Each will do so informed by their own mythologies of AI – the concatenation of different stories and ideologies that shape their expectations and anxieties around what this technology can be. Understanding how AI will develop requires, therefore, an understanding of the many sites in which its story is unfolding*" (p. 5).

Infine col terzo caso, quello 'storico-concettuale', il focus della ricostruzione storica si centra sulle scienze e sulle pratiche di comprensione (e compressione) dell'umano attraverso teorie e discipline riduzioniste dello stesso. La prospettiva individuata (Donnelly 2024) è che:

> "(…) *the idea of artificial intelligence emerged out of reductive and simplistic methodologies pursued in the collective fields of knowledge generally known as the social sciences, or what were for centuries called the sciences of man. Going back four centuries to the idea of "mechanical philosophy" – which attempted to describe nature through machine analogies – the book covers a number of approaches that sought to study human thought and behavior using the assumptions, tools, and techniques of the natural sciences*" (pp. 3-4).

Dunque, in linea con la nostra prospettiva, il dibattito storiografico contemporaneo sull'AI si arricchisce oggi di altre viste (sulle nuove divisioni del lavoro, sulle emergenti diversità nella cultura, sulle scienze riduzionistiche dell'umano) che rappresentano un significativo avanzamento teoretico nella nostra conoscenza non meramente tecnologica dell'ingegneria dell'Intelligenza artificiale. E non solo. Da ultimo, dobbiamo aggiungere qui – con la visione evolutiva del filosofo della tecnologia Benjamin Bratton – anche un'interpretazione dell'Intelligenza artificiale come stadio di sviluppo dell'intelligenza planetaria più complessiva. Lo anticipavamo nell'introduzione parlando di astrofisica. Il progetto e la rete di ricerca *Antikythera* che porta avanti questo programma di ricerca ne è una chiara testimonianza ai nostri giorni[1]. Questo rapido excursus intende far notare illustrativamente (non certo esaustivamente) come anche la ricerca storica contemporanea abbia percepito con più forza la necessità di sollevare lo sguardo ricostruttivo oltre la connotazione strumentale per osservare al meglio (più culturalmente, socialmente e politicamente) la questione complessa dell'AI. Infine, prima di entrare nel merito della proposta e della discussione intorno ai tre approcci, è utile fare un passaggio su definizioni concettuali e operazionali. Al di là, infatti, delle diverse connotazioni e qualificazioni storicamente succedutesi, oggi la comunità giuridica e legale nel nostro Paese (dentro quella più complessiva e competitiva del Pianeta) ha adottato delle definizioni

1 Si veda *Antikythera Studio Report,* https://view.publitas.com/nicolay-boyadjiev/2023-antikythera-studio-report/page/1.

di 'diritto' stringenti e vincolanti per organizzazioni e imprese che sono impegnate nell'ideazione, nella realizzazione, nella commercializzazione di applicazioni e dispositivi basati sull'Intelligenza artificiale. È naturalmente d'obbligo fare qui un accenno al contesto originario della denominazione dell'AI. Riprendiamo dal documento di proposta del seminario di Dartmouth (McCarthy *et al*. 1955) le prime righe che sintetizzano l'obiettivo dello studio estivo ipotizzato all'epoca. Scrivono i proponenti:

> "*We propose that a 2-month, 10-man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves*" (p. 1).

Di questo incipit è opportuno mettere in rilievo la congettura che muove lo studio. Vale a dire, come chiariscono i proponenti, che l'apprendimento e ogni altra caratteristica dell'intelligenza possa essere in principio descritta così precisamente al punto che si possa costruire una macchina in grado di simularli (l'apprendimento umano e l'intelligenza umana). Lo sottolineiamo perché, come avremo modo di argomentare più avanti, questo presupposto con l'arrivo del *machine learning* e del *deep learning* non risulta più dato e necessario (le macchine 'imparano' senza necessità di descrizione). E questo fatto ha implicazioni notevolissime (ancorché non chiare ancora a molti) sulle questioni del confronto umani-macchine. Ad esempio, per quanto riguarda la polarizzazione dei task nella composizione del lavoro (*task/job*) e nell'organizzazione del lavoro (*user/agent*). Con un salto da quella storica ipotesi di ricerca all'attuale impianto regolatorio, la definizione di Intelligenza artificiale si circoscrive oggi operazionalmente alla seguente descrizione. L'articolo 3 (EU AI ACT 2024) recita: "*AI system means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments*".

Tenendo sullo sfondo questa articolazione definitoria, propongo di suddividere le riflessioni intorno alla questione dell'Intelligenza artificiale e dei suoi impatti su economia e società su tre livelli (distinti, ma certamente interconnessi): strumentale, strutturale, istituzionale. Questa mia stratificazione interpretativa servirà a chiarire gli orizzonti concettuali e strategici con cui occorre confrontarsi per approntare un'analisi (più) sofisticata dell'evoluzione tecnologica in divenire.

2. L'approccio 'strumentale' all'AI

La prospettiva strumentale intorno all'AI è oggi quella maggiormente diffusa e più comunemente accettata. Le analisi strumentali si fondano eminentemente su una matrice logica e interpretativa fondamentalmente ingegneristica. L'Intelligenza artificiale viene così ridotta a mero artefatto da scomporre e valutare nelle sue componenti tecniche, computazionali e algoritmiche. In filosofia della tecnologia l'orientamento 'strumentalista' è quello che considera la tecnologia uno strumento che l'essere umano usa come mezzo per un fine (Nyholm 2023). La gran parte delle analisi e delle elaborazioni correnti ha questa prospettiva di fondo che viene adoperata di norma anche, ad esempio, per valutare se un certo 'modello' o 'algoritmo' o 'dispositivo' o 'robot'o 'macchina' sia in sé e per sé, di volta in volta, artificialmente 'intelligente', 'senziente', 'cosciente' e così via. Si accompagna, in genere, anche ad un processo di 'reificazione' dell'Intelligenza artificiale (o anche viceversa di 'personificazione', quando antropomorfizzata). Anche la pratica dei test di benchmark con cui si confrontano gli avanzamenti delle ingegnerie dell'Intelligenza artificiale con l'intelligenza umana ha questo orizzonte interpretativo di riferimento. Questo riduzionismo strumentalista 'cosifica' inconsapevolmente e ingenuamente quello che in realtà sarebbe da considerare piuttosto come un assemblaggio socio-tecnico complesso (Crawford 2022). È evidente che il passaggio dal pensare per artefatti strumentali al pensare per sistemi socio-tecnici non è (e non sarà) per la nostra civiltà uno slittamento ermeneutico semplice e rapido. La nostra tradizione culturale (a partire da quella occidentale moderna) è sicuramente più familiare con un'ontologia solipsistica fatta di

cose, oggetti, strumenti chiusi in sé. E poi, ancora, di meccaniche, di meccanismi e di macchine dentro un processo di continua e invisibile reificazione della realtà. Ma così facendo la nostra civiltà (occidentale) fatica a interpretare in profondità il mondo artificiale in costruzione e le sue articolazioni socio-tecniche. Complessità fatte, piuttosto, di interazioni, di trame, di corrispondenze, di disposizioni, di sovrapposizioni, di stratificazioni, di percolazioni, di ridondanze, di emergenze, di interferenze, di perturbazioni. Non di 'cose', quindi. Prendiamo, per fare un caso, l'attuale evoluzione dell'AI generativa nella forma dei *large language model* (modelli linguistici su larga scala). La manualistica corrente li descrive tecnicamente come dei *text predictor*, dei predittori probabilistici di parole. Il che è strumentalmente vero, come ci insegna la teoria e la pratica del processamento macchinico del linguaggio naturale umano (NPL). In effetti secondo Kublik e Saboo (2022):

> "*Language modeling is the task of assigning a probability to sentences in a language. Simple language models can look at a word and predict the next word (or words) most likely to follow it, based on statistical analysis of existing text sequences. Besides assigning a probability to each sequence of words, language models also assign a probability that a given word will follow a certain sequence of words. In order to create a language model that successfully predicts word sequences, you need to train it on large sets of data. Language models are a key component in natural language processing applications. You can think of them as statistical prediction machines, where you give text as input and get a prediction as the output*" (p. 4).

Rimanendo confinati nella pur giusta (ma limitante) interpretazione strumentale dell'intelligenza generativa di linguaggi sintetici, siamo certamente in grado di cogliere le tecnicalità alla base delle evoluzioni tecnologiche. Ma così facendo non si evoca la complessità e le implicazioni dell'operazione socio-tecnica che produce un linguaggio sintetico come quello generato dai modelli linguistici. Operazione che coinvolge molte delle dimensioni di un atto linguistico, da quelle semantiche a quelle semiotiche e molte altre ancora (Kockelman 2024). Così:

> "(…) *at a certain level of abstraction, a large language model may be understood as a parameter-dependent function that accepts a sequence of words as its input and returns a sequence of words as its output. Assuming the function was well chosen and its parameters have been adequately set, the inputted sequence, known as the prompt, specifies a task that the user wants fulfilled, and the outputted sequence, known as the response, fulfils that task (…) More carefully, the prompt is typically a sequence of words that describes, and perhaps demonstrates, certain satisfaction conditions, and the response – at least when all goes well – is a sequence of words that satisfies those conditions. Phrased another way, if prompts are descriptions of actions that the user wants the model to undertake, responses are the results of the actions so undertaken. In short, at this level of abstraction large language models are very complicated semiotic agents, with prompts as their signs, responses as their interpretants, satisfaction conditions as their objects, and parameters rather than values as their guiding principles*" (p. 29).

Se ci limitiamo alla dimensione strumentale non comprendiamo la radicalità socio-tecnica del passaggio a questa parola 'non-umana'. Infatti, questa presa di parola della macchina sta facendo collassare, d'un colpo, il senso del linguaggio naturale – orale, scritto e poi stampato – e della scrittura umana per come si è sin qui storicamente evoluto. Con questo cedimento si incrinano anche altri domini di senso implicati: l'autorialità, la normatività, la responsabilità, l'autenticità umane. Insieme ad essi, più radicalmente, è anche l'erosione della civiltà dello scrivente e del parlante, dell'autore e del lettore, dell'ordine del discorso e dei regimi di verità creati su questo nel corso del tempo (Accoto 2024). E ancor di più, direi. Questa trasformazione ingegneristica non riguarda solo la capacità sorprendente di processare macchinicamente il linguaggio naturale umano. Più complessivamente riguarda e ha impatti significativi anche sulla natura del lavoro cognitivo perché scardina l'implicito principio cautelativo contenuto nell'incipit della proposta del 1955. Infatti, a quasi settant'anni da quella proposta, discutendo oggi di lavoro e Intelligenza artificiale, il punto su cui va portata l'attenzione è proprio quello relativo alla congettura della descrizione precisa di apprendimento e intelligenza. Se c'è o è possibile questa descrizione,

allora è ipotizzabile e progettabile una macchina che simuli le capacità cognitive umane. Se non c'è, la macchina non sarà in grado di replicare quel compito. Negli anni, questo vincolo o limite nella nostra capacità di descrivere come siamo in grado di fare certe cose (es. riconoscere un oggetto o scrivere un testo) ci ha messo al riparo dalla piena automazione. Non sapendo descrivere come facciamo umanamente a fare certe cose, non saremmo riusciti a operazionalizzare e codificare molte attività e pratiche. Perché, come insegna il principio di Polanyi (conoscenza tacita), noi conosciamo molto più di quello che riusciamo a descrivere. Questo principio è stato a fondamento dei modelli economici impiegati per delineare gli orizzonti del lavoro umano a fronte dell'arrivo della conoscenza macchinica. Tuttavia, l'arrivo di un'Intelligenza artificiale in grado di 'imparare dall'esperienza' come si dice comunemente (*machine* e *deep learning* anche nel caso del linguaggio processato dalle macchine), di conoscere il mondo senza che noi che la istruiamo sul da farsi, cambia di nuovo l'orizzonte del lavoro umano. Il vincolo cautelativo del principio di Polanyi viene a decadere e, anzi, proprio si inverte (Autor 2014). Non solo la macchina è in grado di fare (simulare) cose senza che noi le codifichiamo prima (tipo il parlare, il vedere, lo scrivere), ma come fa a farle ci è oscuro in buona misura (*black box*). Fa cose che noi non sappiano precisamente come fa a fare. Di fronte a questa torsione epistemica (di conoscenza) ed ergodica (di lavoro), i modelli con cui avevamo cominciato a sondare il rapporto tra lavoro umano e Intelligenza artificiale vanno in fibrillazione. Prima si valutava la capacità della macchina di svolgere compiti routinari codificabili lasciando a noi quelli complessi (come il linguaggio), non traducibili in operazioni computazionali. Ora quel modello per *task polarization* scricchiola (Autor 2014) e le recenti sicurezze di essere al riparo dall'automazione si mutano in imponderabili incertezze.

3. L'approccio 'strutturale' all'AI

In questa tecnomorfosi, non solo il lavoro in quanto tale, ma anche l'impresa si sta trasformando e sta perdendo i suoi connotati storici, la sua struttura tradizionale e con essa l'organizzazione del lavoro. Per questo, sollecitato dalle necessità produttive e competitive contemporanee, il fenomeno dell'Intelligenza artificiale ha cominciato a essere oggetto di uno studio più strutturale da parte dell'analisi economica (e non solo strumentale). Perché è chiaro che per le imprese questi prossimi mesi e anni saranno cruciali per capire se e come riusciranno a trarre valore complessivo dall'introduzione delle ingegnerie dell'Intelligenza artificiale nei loro processi di produzione. Ma per allontanarsi dall'isteria della cronaca e focalizzarsi su una strategia più ponderata, sarà necessario dotarsi di nuovo pensiero e nuove analisi. E, tuttavia, difettano oggi le riflessioni ponderate improntate alla teoria e alla strumentazione economica. Di certo per le imprese (e le istituzioni) è necessario arginare l'incertezza che circonda questa nuova ondata tecnologica che rimane in un'evoluzione caotica e sconosciuta, un po' differente da altre storie di tecnologie d'impatto vissute in passato (Naudé *et al.* 2024). Al contempo è anche oramai chiaro che anche la teoria economica dovrà adeguare e rinnovare il proprio patrimonio metodologico e teoretico per fronteggiare l'arrivo di una tecnologia profondamente trasformativa di equilibri, dinamiche e assetti. In questo senso:

> "(…) *the field of economics also needs to adjust its tools to be able to illuminate AI better. Key models in economics, for example, growth models, have until recently wholly abstracted from technology (and energy), focusing only on the nineteenth-century world of capital and labor as production factors. It was only in the 1990s that technology was endogenized, and the key feature of technology – as ideas that offer increasing returns and combinatorial possibilities– incorporated into economic growth models. These insights earned Paul Romer a Nobel Prize. It would be premature to imagine that the Information and communication technology (ICT) revolution, which has gathered speed only after 2007, is adequately reflected in these models. We need more details and realism of digital technologies, and specifically AI, to be included in our models. These will offer gains in understanding how and why, and when, AI affects key economic outcomes such as economic growth, inequality, productivity growth, poverty, innovation and investment rates, wages, and consumption*" (p. 5).

L'Intelligenza artificiale non è, allora, semplicemente un nuovo prodotto o servizio, come ingenuamente si continua per lo più a ritenere. Piuttosto è nuova fabbrica (Accoto 2024), vale a dire nuova architettura produttiva per orchestrare

in modi altri le interazioni socioeconomiche umane e non-umane. Riuscire a passare da un approccio strumentale (o per 'soluzioni') ad un approccio strutturale (per 'sistemi') richiede, tuttavia, una visione strategica che non molte imprese e organizzazioni, leader e professionisti riescono oggi ancora a sviluppare. D'altro canto, è accaduta già in passato in altri momenti e per altre rivoluzioni tecnologiche questa difficoltà cognitiva nel valutare il cambiamento tecnologico in mondo 'sistemico' e non solo 'soluzionista'. Se teniamo per valida l'analogia dell'introduzione presente dell'Intelligenza artificiale nelle organizzazioni con l'introduzione storica dell'elettricità nelle fabbriche (Agrawal *et al.* 2022), potrebbe chiarirsi meglio questo passaggio dalla soluzione al sistema.

> "(…) *Some value propositions are more attractive than others. In the case of electricity, point solutions and application solutions predicated on directly replacing steam with electricity without modifying the system offered limited value, which was reflected in industries' slow initial adoption. Over time, some entrepreneurs saw the opportunity to deliver system-level solutions by exploiting the ability of electricity to decouple the machine from the power source in a manner that was impossible or too expensive with steam. In many cases, the value proposition of system-level solutions far exceeded the value from point solutions*" (p. 11).

Le soluzioni verticali e applicative (strumenti) una volta inquadrate più strategicamente a livello di 'sistema' scardinano gli assetti industriali e le architetture produttive. L'adozione dell'Intelligenza artificiale sta avvenendo oggi più spesso per introduzione destrutturata e tattica di 'strumenti': algoritmi di ottimizzazione, applicazioni per efficientare i flussi, servizi di personalizzazione e predizione e così via. Choudary li definisce "effetti del primo ordine" dell'AI (Choudary 2025 *forthcoming*). Tuttavia, questo è solo lo stadio iniziale della trasformazione come lo fu all'epoca la prima fase di adozione dell'elettricità nelle fabbriche. Una volta acquisito questo primo momento (strumentale) più orientato all'automazione e all'efficientamento (con riduzione dei costi ad es.), l'innesco per una più radicale morfosi e ridisegno strutturale dei processi si avvierà con gli "effetti di secondo ordine", più sistemici diremmo.

> "*The dominant approach to analyzing AI today is as a technological tool. This comes with all the mental models of tools and technologies – we look at AI through the lens of mechanization, automation, and optimization. While AI does perform these functions, these mental models are insufficient for understanding its true impact – they limit us to the first order effects of AI. Reshuffle introduces a systems perspective, arguing that AI should be understood primarily as a coordination mechanism, rather than merely a tool for automation. Automation affects tasks. Coordination reshapes entire workflows, organizations, and economies. To see AÌs full impact, we need to explore its second-order effects, particularly how it restructures knowledge-based work and the systems within which work happens*" (Choudary, *blogpost*, marzo 2025).

In questa prospettiva, anche l'idea comune che non sarà l'AI a rubarti il lavoro, ma un umano che userà l'AI è una visione ingenua del cambiamento strutturale in atto. Perché l'impresa-piattaforma (o *platfirm* come l'ho rinominata con un neologismo) con i suoi cicli iterati di addestramento macchinico che impara dall'esperienza (e cioè, nel nostro caso, dal lavoro cognitivo umano con cui coevolve in molti casi parassiticamente ad oggi) costringe costantemente a riprogrammare le relazioni *task/job* e *user/agent*. Dunque, per stare al nostro esempio sull'Intelligenza artificiale generativa, un motore linguistico su larga scala non sarà solo un *text predictor*, ma diviene un *knowledge engine* in grado di trasformare radicalmente i processi produttivi aziendali, i ruoli e le competenze delle risorse umane, le logiche di business e le dinamiche di creazione di valore. Non si tratterà allora solo di riduzione dei costi che un sistema linguistico automatizzato riuscirà a produrre, ma di ridisegno più complessivo delle architetture informative e delle infrastrutture cognitive di organizzazioni e imprese. E anche oltre queste.

4. L'approccio 'istituzionale' all'AI

Ancora più marginale nella riflessione odierna sugli impatti dell'Intelligenza artificiale è la prospettiva 'istituzionale' che guarda all'AI. Tra diritto, economia e politica, l'orientamento istituzionale cerca di allargare la riflessione dagli aspetti 'instrumentanti' e 'infrastrutturanti' a quelli

'istituzionalizzanti'. Non guarda quindi solo alle ingegnerie delle soluzioni o alle architetture dei sistemi, ma esplora più radicalmente la capacità 'istituente' della tecnica. In questa terza vista, non è solo la singola composizione del lavoro a essere osservata nelle trasformazioni innescate dall'AI, né solo è la specifica struttura dell'impresa con le sue nuove organizzazioni del lavoro ri-orientate dall'AI. Bensì è la profonda metamorfosi di istituzioni come i mercati, le burocrazie, gli stati. È il loro progressivo scardinamento ed erosione in ragione dell'emergere di nuove forme istituzionali di coordinamento, di comunicazione, di decisione, di coercizione, di informazione dentro future civilizzazioni umane. Pensiamo, di nuovo, ai modelli linguistici su larga scala oggi molto in voga. Possiamo analizzarli semplicemente come 'strumenti' (macchine calcolatrici delle parole), come text predictor, quindi soluzioni probabilistiche che predicono la parola successiva data una sequenza di parole. E questo è certamente vero, ripetiamolo, da un punto di vista ingegneristico. Poi possiamo anche interpretarli come nuove 'sistemi' di gestione dell'informazione, come emergenti applicazioni per architetture informative aziendali utili a ottimizzare e performare la circolazione delle informazioni (motori generativi di conoscenza) oppure ancora di dialogo e interazione esterna con consumatori e clienti. E questo è anch'esso vero, naturalmente, da un punto di vista economico. Oppure possiamo, infine, iniziare a considerarli come nuove (forme di) istituzionalità socio-politiche similmente ad altre più storiche istituzioni umane come i mercati (Farrell *et al*. 2025). In questa prospettiva

> "(…) *large models combine the features of cultural and social technologies in a new way. They generate summaries of unmanageably large and complex bodies of human generated information. But these systems do not merely summarize this information, like library catalogs, internet search, and Wikipedia. They also can reorganize and reconstruct representations or "simulations" of this information at scale and in new ways, like markets, states, and bureaucracies. Just as market prices are lossy representations of the underlying allocations and uses of resources, and government statistics and bureaucratic categories imperfectly represent the characteristics of underlying populations, so too are large models "lossy JPEGs" of the data corpora on which they have been trained*" (p. 1154).

Certamente può essere straniante considerare i modelli linguistici su larga scala come istituzionalità, ma non dovremmo poi stupirci più di tanto. Anche i mercati economici hanno la stessa capacità istituzionale di processare l'informazione collettiva (e anche loro unendo intelligenza e stupidità) per determinare azioni e decisioni collettive (Farrell *et al*. 2025). Un secondo, ulteriore elemento da considerare dentro un approccio che possiamo definire come 'neoistituzionale' è l'arrivo degli agenti autonomi dentro la società (oltre che dentro l'economia). Nello sviluppo dell'Intelligenza artificiale, la fase successiva all'emergenza dei modelli linguistici su larga scala sono infatti gli agenti artificiali autonomi. In effetti, l'era degli *artificial autonomous agents* avvia la definitiva trasformazione delle macchine da meri strumenti di produzione a sofisticati attori economici, sia come macchine consumatrici di servizi sia come macchine creatrici di imprese. Una volta dotate di autonomia operativo-cognitiva (*ai agents*) e da ultimo anche operativo-finanziaria (*crypto wallets*), le macchine con 'menti artificiali' e 'portafogli digitali' trasformeranno radicalmente la natura istituzionale dei mercati. Diverranno di fatto dei *market maker,* dei creatori di nuove forme di mercati. E occorrerà anche valutare se 'mercato' sarà a quel punto ancora il concetto giusto per significare istituzioni che vivranno di vita nuova. Tra assettizzazione del codice e agentificazione dell'economia, prosegue così il percorso sorprendente e arrischiato dell'economia della macchina. Questa *machine economy* è sicuramente uno dei modelli e delle strategie emergenti in grado di trasformare le tradizionali logiche di mercato e le classiche forme dell'impresa. Se, dunque, il tema filosofico dei nostri giorni è l'agentività della macchina (Mattingly e Cibralic 2025), anche sul fronte 'istituente' i percorsi trasformativi in questa direzione si sono avviati. Leggere questi inattesi passaggi tecnologici nella loro dimensione istitutiva (cioè come nuove istituzioni umane), come abbiamo visto, e non solo come un passaggio tecnologico, è un compito strategico e prefigurativo vitale per le organizzazioni e le imprese che intendono prosperare dentro gli orizzonti della nuova economia artificiale tra *agents* e *wallets*. Naturalmente abbiamo ancora un lungo cammino da percorrere per trasformare tutto il 'lavoro' in

'codice' e passare così dalle attuali *chain-of-thought* alle future *chain-of-work* anche con capacità di spesa autonoma delle intelligenze artificiali (*ai agent* + *crypto wallet*). Ma la direzione è oggi segnata e prevede di passare dall'attuale *robotic process automation* (RPA) alla *agent process automation* (APA) istituzionalmente trasformando logiche di mercato, dinamiche di coordinamento, meccanismi e agenti decisionali (Pinol 2025). Nell'era degli ecosistemi a piattaforma, la cocreazione di valore delle imprese si riconfigurerà sempre più istituzionalmente come un processo 'catallattico' (di scambio o transazionale), 'simpoietico' (di coevoluzione o coopetitivo) e 'prolettico' (di anticipazione o feedforward) con cui attori socioeconomici eterogenei – antropici e non antropici – automaticamente e contestualmente scambieranno servizi-per-servizi integrando risorse di varia natura (operanti e operande, tangibili quanto intangibili, proprietarie e non proprietarie, automate ed eteromate, algoritmiche e protocologiche, umane e agentiche). La questione della creazione del valore (e del senso) e della sua conservazione e circolazione dentro questi stack istituenti potenziati dall'Intelligenza artificiale (ricognitiva, predittiva, generativa, agentiva) dovrà dunque essere nuovamente ripensata alla luce delle stratificazioni esplorate: strumentali, strutturali e istituenti. Come abbiamo mostrato nel caso di specie degli LLM, riuscire a leggere questo passaggio dei *large language model* dall'essere *text predictor* al divenire *knowledge engine* a rappresentare dei *market maker* è la sfida che oggi società ed economia fronteggiano.

Conclusione

Dunque, sarà necessario che il nostro pensare intorno alle nuove ingegnerie (intelligenze artificiali, ma pensiamo anche ad es. alle reti decentralizzate *blockchain-like*) evolva oltre la dimensione '(in) strumentale' – con cui ingenuamente molti le analizzano ancora – per andare verso una loro interpretazione più '(infra)strutturale' e finanche '(neo) istituzionale'. È strategico cominciare a pensare, ad esempio, ai modelli linguistici su larga scala (ma anche ad es. ai registri decentralizzati a consenso crittografico) non come mere strumentalità tecnologiche (dagli algoritmi di back-propagation ai protocolli di zero-knowledge proof ai dispositivi sim-to-real/real-to-sim alle reti di networking quantistico e altro ancora) per 'soluzioni' e 'sistemi', ma come emergenti forme di 'istituzioni' socioeconomiche. E cioè potenziali (certamente anche arrischiate) istituzionalità *più-che-umane* (umane e robotiche) che si affiancheranno e si incroceranno (e col tempo anche in competizione direi come già si inizia a intravedere) con altre, diverse e più antiche istituzioni come le imprese, i mercati, le burocrazie, gli stati. Non si tratta allora solo di tecnologie *general-purpose*, ma proprio di *institutional technologies* (Allen *et al.* 2025). Per cogliere speculativamente questa sorprendente capacità 'istituente' della tecnica, dovremo allora dotarci di nuove mappe interpretative, di nuovi significati e nuovi saperi. E questo ci porta ad uno snodo nevralgico di tutto il discorso. Noi siamo oggi impegnati, consapevolmente o meno, in un cruciale momento di interrogazione e sperimentazione sul futuro dell'umano e della sua civilizzazione. A questa sfida epocale (non episodica) stiamo oggi faticosamente tentando di rispondere con strategie varie, auspicate come convergenti: istruzione educativa, regolazione giuridica, conformazione etica, direzione politica. Sono tutte necessarie, naturalmente. Temiamo però anche non sufficienti. Imprescindibili, dunque, ma non bastanti per la missione ardua che ci attende sin da subito. Perché la dimensione ingegneristica in divenire che abbiamo qualificato come 'Intelligenza artificiale' è in ultima istanza *una planetaria provocazione di senso*. Cioè, un radicale attacco culturale all'idea (nel tempo costruita e diveniente) di umano e di civiltà umana, di economia e di politica con i suoi storici strumenti, sistemi e istituzioni. Dunque, con il dispiegarsi planetario dell'Intelligenza artificiale, noi non affronteremo solo 'problemi tecnici' (con vulnerabilità e rischi reali di discriminazioni, manipolazioni, deprivazioni, polarizzazioni, alienazioni, contraffazioni). Noi affronteremo delle 'provocazioni intellettuali'. E se ai problemi tecnici lavoreremo, nel tempo e a tentativi, per trovare una *soluzione ingegneristica* di qualche tipo (informatica, legale, politica e così via), alle provocazioni intellettuali dovremo invece rispondere, di necessità, con l'*innovazione culturale*. Un'innovazione culturale che dovrà immaginare, come ho cercato di argomentare succintamente, rinnovate soluzioni, sistemi e istituzioni 'artificialmente intelligenti'. Con una domanda valoriale di senso non solo istituzionale ma esistenziale: *perché* (lo facciamo) e *per chi* (lo facciamo).

Bibliografia

Accoto C. (2024), *Il Pianeta Latente. Provocazioni della tecnica, innovazioni della cultura*, Milano, Egea

Agrawal A., Gans J., Goldfarb A. (2022), *Power and Prediction. The Disruptive Economics of Artificial Intelligence*, Boston, Harvard Business Review Press

Allen D., Berg C., Potts J. (2025), *Institutional Acceleration. The Consequences of Technological Change in A Digital Economy*, Cambridge, Cambridge University Press

Autor D. (2014), *Polanyi's Paradox and the Shape of Employment Growth*, Working Paper n.20485, Cambridge, National Bureau of economic Research

Bates D.W. (2024), *An Artificial History of Natural Intelligence. Thinking with Machines from Descartes to Digital Age*, Chicago, Chicago University Press

Bratton B. (2019), *The Terraforming*, Moscow, Strelka Institute

Cave S., Dihal K. (2023), *Imagining AI. How the World See Intelligent Machines*, Cambridge, Cambridge University Press

Choudary S.P. (2025), *Reshuffle. Who Wins When AI Restack the Knowledge Economy*, New York, Future First Books (*forthcoming*)

Crawford K. (2022), *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, New Haven, Yale University Press

Donnelly K.P. (2024), *The Descent of Artificial Intelligence. A Deep History of an Idea of Hundred Years in the Making*, Pittsburgh, University of Pittsburgh Press

European Commission (2024), EU Artificial Intelligence Act, *OJ Official Journal* (July 12, 2024)

Farrell H., Gopnik A., Shalizi C., Evans J. (2025), Large AI Models Are Cultural and Social Technologies, *Science*, March 13

Frank A., Grinspoon D., Walker S. (2022), Intelligence as a planetary scale process, *International Journal of Astrobiology*, 21, n.2, pp.47-61

Kockelman P. (2024), *Last Words. Large Language Models and the AI Apocalypse*, New York, Prickly Paradigm Press

Kublik S., Saboo S. (2022), *GPT-3. Building Innovative NLP Products Using Large Language Models*, Sebastopol, O'Reilly

Mattingly J., Cibralic B. (2025), *Machine Agency*, Cambridge, The MIT Press

McCarthy J., Minsky M.L., Rochester N., Shannon C.E. (1955), A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, *AI Magazine,* 27, n.4 <https://doi.org/10.1609/aimag.v27i4.1904>

Naudé W., Gries T., Dimitri N. (2024), *Artificial Intelligence. Economic Perspectives and Models*, Cambridge, Cambridge University Press

Nilsson N.J. (2010), *The Quest for Artificial Intelligence. A History of Ideas and Achievements*, Cambridge, Cambridge University Press

Nyholm S. (2023), *This is Technology Ethics. An Introduction*, Hoboken, Wiley

Pasquinelli M. (2023), *The Eye of The Master. A Social History of Artificial Intelligence*, London, Verso Books

Pinol A. (2025), What It Really Takes to Make AI Agents Work, *NfX blogpost*, February 2025

Cosimo Accoto
cosimo.accoto@unimore.it
Filosofo tech, ricercatore affiliato (MIT Boston), professore aggiunto (Università degli Studi di Modena e Reggio Emilia). Pubblica su diverse riviste tra cui *Economia & Management* (SDA Bocconi School of Management), *Harvard Business Review Italia*, *Sistemi & Impresa*, *Aspenia*, *MIT Sloan Management Review Italia*, *Civiltà delle Macchine*. Autore della trilogia filosofica sulla civiltà digitale: *Il mondo in sintesi* (2022), *Il mondo ex machina* (2019), *Il mondo dato* (2017) tradotto quest'ultimo in più lingue, recentemente anche in cinese. Il suo ultimo saggio è: *Il Pianeta Latente. Provocazioni della tecnica, innovazioni della cultura* (2024). Attuali interessi di ricerca: filosofia del codice e della programmazione, sensor e software society, automazione robotica e teorie dell'Intelligenza artificiale, blockchain e filosofia dei criptosistemi e dei cryptoasset, computazione quantistica e biologia sintetica, filosofia delle realtà estese, sintetiche e immersive.

Artificial intelligence between Big Science and Small Science: an Italian strategy to compete

Luciano Pietronero
Enrico Fermi Research Center CREF, Rome

Marco Trombetti
Translated srl, Rome

This article presents a strategic analysis on the situation and the possible developments of Artificial intelligence (AI) from a scientific but also an economic and business point of view. It contrasts the two complementary paths of Big Science and Small Science by analyzing two case studies: machine translation and the Economic Fitness and Complexity framework, a new methodology that provides superior insight into economic development with fewer but carefully chosen data. It then examine the implications of AI for the labor market and assess Italy's position, which can become a global leader in Small Science for AI by leveraging interdisciplinarity and complexity science, ingenuity, and talent attraction.

Questo articolo presenta un'analisi strategica della situazione e dei possibili sviluppi dell'Intelligenza artificiale (IA) da un punto di vista scientifico, ma anche economico e aziendale. Si confrontano i due percorsi complementari di Big Science e Small Science analizzando due casi di studio: la traduzione automatica e il framework Economic Fitness and Complexity, una nuova metodologia che fornisce una visione approfondita dello sviluppo economico con un numero inferiore di dati, ma accuratamente selezionati. Si esaminano quindi le implicazioni dell'IA per il mercato del lavoro e si valuta la posizione dell'Italia, che può orientarsi a diventare un leader globale nella Small Science per l'Intelligenza artificiale, sfruttando interdisciplinarità e scienza della complessità, ingegno e attrazione dei talenti.

DOI: 10.53223/Sinappsi_2025-02-2

Citazione
Pietronero L., Trombetti M. (2025), Artificial intelligence beetween Big Science and Small Science: an Italian strategy to compete, Sinappsi, XV, n.2, pp.18-30

Keywords
Competitiveness
Artificial intelligence
Labour market

Parole chiave
Competitività
Intelligenza artificiale
Mercato del lavoro

1. Introduction to Artificial intelligence

Fantastic expectations but also existential anxieties are two of the conflicting feelings that often arise in the general confusion of discussions on AI. The aim of this article is to provide a concrete and understandable picture of the AI situation and its possible developments, in particular focusing on how to make Italy a leader in AI. Let's start with the two examples concerning the authors. On the one hand, we have the case of a scientific and entrepreneurial activity in the field of machine translation. This requires the use of large neural networks and the input of a huge amount of data, in the spirit of ChatGPT and similar methods. On the other hand,

Contributo su invito non sottoposto a peer review.

we have the problem of making economic analyses of countries'growth more scientific and accurate. This requires a careful selection of a few selected data and the introduction of a new dedicated algorithm from which a better analysis of the growth of countries is obtained than that of large traditional economic institutions. Two very different and apparently conflicting cases with respect to the use of data, many or few, from which, however, one can extrapolate a varied and fascinating overall picture, of which we have only seen the beginning. A joint hybrid scheme then makes it possible to assess the impact of AI on the labor market based on actual investments rather than opinions. Finally, we will try to draw precise and realistic conclusions on how to best use the opportunities of this epochal revolution for Italy.

A bit of AI history and the invention of the wheel

Talking about Artificial intelligence immediately evokes an analogy with the human brain. Over the past fifty years, neurophysiology has allowed us to develop a detailed knowledge of the brain from a morphological point of view. This, however, has not yet led to a complete understanding of its functioning, which in many respects is still a mystery. It is a very complex network whose sites are represented by neuronal cells or neurons and the links are the synapses. In a brain there are about 100 billion neurons and 150 trillion synapses. The number of synapses per neuron can vary from 5,000 to 100,000, an incredibly high figure that gives an idea of the enormous complexity of the entire network. How intelligent thought can arise from this raw material is one of the most fascinating mysteries.

Compared to the specific properties of individual cells, great progress has been made, but the vision of the collective functioning of the system remains mysterious in many respects. Neurons behave a bit like switches that can be activated or be inert. When a neuron is activated, through an appropriate electrical stimulus, it generates an electrical signal that is transmitted to the synapse. This releases chemical messengers, neurotransmitters, which activate the receptors of the other connected neuron. This signal is converted into an electrical impulse and, through the synaptic network, can activate other neurons. These signals can be stimulators or inhibitors. A neuron generates a weighted sum of incoming binary signals coming from other neurons to generate an output binary signal.

An important element was then to realize that memory in the brain is associative in nature. In traditional computers, memory is addressed in specific places with a unique address, while in the brain even partial information allows you to get to the right concept. For over fifty years, we have tried to mathematize the behavior of neurons to try to artificially reproduce brain activity, but it was only in the 1980s with the work of Hopfield and Hinton (The Nobel Committee for Physics 2024) that a qualitative leap and the development of modern Neural Networks took place. This associative nature of memory, together with the binary functioning of neurons and synaptic connections, have inspired analogies with the models of Statistical Physics and the related concepts of energy, temperature and entropy. In particular, the complex energy structures of these models made of valleys and mountains have provided a possible interpretation of associative memory. If a given configuration of the entire system corresponds to a precise concept, entering an approximate concept into the system corresponds to being close to this energy minimum and, by appropriately relaxing the system, the correct result will be obtained even starting from partial information.

This analogy with Statistical Physics and, in particular, spin glasses (The Nobel Committee for Physics 2021) is at the basis of the introduction of Neural Networks, which can be of various types and correspond to an emulation of the properties of the brain that has associative memory and other important conceptual analogies. Of course, as with the natural brain, a long process of experience and education is necessary, and these networks also need to be trained with appropriate data. All this has led to increasingly interesting developments in recent decades, up to the current explosion of interest in generative AI and Large Language Models (LLM). Of course, trying to artificially reproduce intelligence based on analogies with the human brain is fascinating and certainly very valid. After all, the brain originates from Darwinian evolution, which is generally a very effective mechanism for finding ways to perform certain functions.

But it is also legitimate to ask whether this is necessarily the only method. Many of man's inventions are in fact inspired by nature but not all of them. For example, the wheel is one of the main

inventions of humanity, but it is not present in any natural organism. We can therefore ask ourselves if, in addition to trying to replicate the structure of the brain, we could invent other concepts and structures that have not been explored by Darwinian evolution.

The invention of the wheel is attributed to the Mesopotamian civilization around 3400 BC and then to China. But suppose that these brilliant ancestors of ours, to improve transport, had been inspired by living beings such as people and animals. They would have tried to build robots to replicate their movements but clearly, with the technologies of the time, they would not have gone far. On the contrary, by observing the rolling of the trunks they used them to slide heavy objects such as stones for the pyramids. Then came the remarkable idea of taking only one section and inserting an axis passing through the center and thus the wheel with the central axis was born, one of the most important discoveries of humanity. Why do we mention this example? Because it is often assumed that the only way, or at least the best way, to extract information from a large amount of data is to emulate the human brain. In fact, this is the result of Darwinian evolution with respect to this problem but, as with the wheel, it is not said that it is the only or always the best. It is interesting to note that the wheel is completely absent in the Darwinian evolution of any species, probably because all organisms that come from cells are inevitably connected topologically and cannot have separate parts that are part of the same individual. However, this observation stimulates us to think of different perspectives with respect to the emulation of the brain.

An example of an alternative situation is Google's Page Rank algorithm, introduced in 1996, which allows categorizing all Internet sites with a simple algorithm. The Internet is a complex network in the sense that it is highly heterogeneous, it was not designed by anyone and grows from the bottom with the addition of new sites without central control. Each site contains links to other sites, and the problem was to classify these sites according to their importance. A traditional way could be to study the information contained in each site, compare it, and then make a ranking on this basis. Clearly, this method would be extremely expensive and also subjective, but the main problem is another. If a site contains a lot of information we would be tempted to give it a high ranking, but if no other site points to this site, it means that this information is not considered interesting for the community to which it belongs. So, if we consider the links between the various sites and not the characteristics of the site itself, we have direct information from the experts of the community of reference. This situation poses a technical problem of non-locality in the sense that the connections that are important to evaluate a given site are not the outgoing ones that are easily verifiable, but the incoming ones that potentially depend on any other site. Moreover, the importance of a connection clearly also depends on the importance of the site from which it originates, and all this complex information must be treated in a self-consistent way. This seemingly very complex problem is solved brilliantly by Google's algorithm which is based on a simple process of diffusion on the structure of Internet sites. By examining this process with a stochastic simulation method (Monte Carlo type), it can easily be verified that the self-consistent relevance of a given site is connected to the probability that the diffusion process will transit through that site. In this way we obtain a ranking of all the sites of a certain community (defined by the question that we introduce in the system) in which the information defined by the connections of the sites of the community is exploited, without having to study the content of each site individually. This simple algorithm applied to the structure of the Internet produces a hierarchy of all the sites related to the question asked and has allowed the whole world to explore the network of Internet sites in an extremely effective way, as we can verify every day. For many, Google's Page Rank algorithm represents the true beginning of modern AI.

Interdisciplinarity and Complexity

What lessons do we take from this? First of all, we can ask ourselves which scientific discipline the Google algorithm belongs to. Clearly it is computer science, but also statistical physics and mathematics, and even a bit of biology, considering that the Internet develops with autonomous growth from the bottom. In other words, the Google algorithm, which has revolutionized all our information and produced one of the most important companies in the world, cannot be included in the context of traditional scientific disciplines. But this should not surprise us considering that Physics, Chemistry, Mathematics etc. were defined as such towards the

end of the eighteenth century, after the Napoleonic reforms and the development of Enlightenment. But after three centuries the world is radically different and the scientific and social problems of today's world are completely different and transcend these disciplines. In fact, the Google algorithm was developed by two students in a garage and not in a prestigious academic institution. This should make us reflect on the importance of interdisciplinarity in trying to solve the great current problems and on the fact that overcoming disciplinary boundaries with dedicated institutions can be extremely interesting and important.

Finally, we see that Google is not based on a general Neural network but develops a specific algorithm for a specific problem. An interesting question is whether a suitable neural network, appropriately trained, could generate the same results without explicitly using the Google algorithm. In our metaphor, Google's algorithm goes more towards the invention of the wheel than towards the emulation of the brain.

The fact that these systems focus more on the connections between the various elements of the system and that from this strongly interconnected network emerging properties such as intelligence originate, leads us naturally to the field of Complex Systems Science. The article published in *Europhysics news* (Pietronero 2008) is a sort of Manifesto for the Physics of Complex Systems that inspired the foundation of the homonymous CNR Institute. This article proposes a suggestive metaphor according to which Particle Physics studies the *elementary bricks* of matter while the Physics of Complex Systems studies the *Architecture* of nature, focusing mainly on mutual interrelations and related emerging properties. Clearly this depends on the properties of the bricks, but then it has its own characteristics and fundamental laws that are not attributable to the bricks and require characteristic concepts and methods. It is also clear that this paradigm of Complexity can be naturally extended to interdisciplinary situations in which the bricks are not elementary particles but, for example, people or institutions as in the field of economics. The goal is to make these disciplines more scientific by using concepts and methods from the Physics of Complex Systems. One may object that in some fields it is difficult to do experiments. This is true also in Astrophysics, which is essentially an observational science but studied with the scientific method of Physics. Today, in Economics, we have a huge amount of data and therefore the application of concepts and methods from Physics can lead to an advancement of the scientific level in issues of the highest scientific and social importance such as the economic development of countries.

2. From translation to the future of AI

Machine translation has always been the most complex challenge for Artificial intelligence because language represents the deepest manifestation of human intelligence, rich in nuances and ambiguities that are difficult to replicate by a machine. In the 1940s, Alan Turing laid the theoretical foundations by translating the Enigma code, realizing that decoding messages was equivalent to a form of language translation. In the 1950s, the initial enthusiasm for the first translation machines quickly collided with the enormous complexity of natural language, highlighting insurmountable limitations for decades.

Only with the advent of statistical methods in the 1990s and deep neural networks since 2010 has machine translation improved significantly, culminating in the invention of the Transformer in 2017. This architecture, initially designed to overcome the limitations of language translation, has revolutionized the entire field of AI thanks to its ability to understand and generate texts with almost human quality. Today, Transformers (Vaswani *et al.* 2017), which were created to translate between languages, are the basis of ChatGPT, a system that 'translates' questions into coherent answers, demonstrating how machine translation remains the privileged ground on which to measure the progress of Artificial intelligence.

In recent years, Artificial intelligence has gone through three fundamental stages: initially, neural networks outperformed university students in answering notional questions. Subsequently, models such as ChatGPT have shown superior language and persuasive skills to almost all the same students. Since 2024, research has focused on enhancing logical reasoning skills through techniques such as Chain-of-Thought, with systems already approaching the logical abilities of doctoral students.

The human brain, used by many as the absolute reference of intelligence, is certainly the most sophisticated form we know, but there are no physical limits that prevent the creation of

superior intelligences. However, the limitations of current techniques and data are also beginning to emerge clearly. These limitations represent a huge opportunity for new players, including Italy, to enter the market.

Current techniques are mainly limited in three areas:

- *Data*: today AI systems are mainly trained on textual data, ignoring the enormous potential of multimodal learning (for example, audio and video). An equipped machine, capable of processing from sensors superior to human ones, could develop innovative and more efficient learning methods.
- *Algorithms*: the current Transformer (Vaswani *et al.* 2017) architecture has limited learning and efficiency capabilities and lacks plasticity, i.e. the ability to modify its architecture, without human intervention, to adapt to evolving contexts of use. It is still 2-3 orders of magnitude more expensive in data and energy than the presumed practical limits and 5 orders of magnitude more expensive than Landauer's theoretical limit.
- *Computing*: the AI chip market (GPU) today has a strong monopoly, with economic barriers to entry, ecosystem constraints and high technological lock-in. Stimulating the creation and adoption of new chips could significantly expand access to innovation in AI and foster a more competitive and diversified ecosystem.

Taking advantage of these opportunities can position Italy as a key player in the development of the Artificial intelligence of the future.

3. Economic Fitness Data Science Paradigm for countries'development

Economic Fitness and Complexity (Pietronero *et al.* 2024) (EFC) is the recent economic discipline and methodology initially developed at Sapienza University of Rome and ISC-CNR and now at Centro Fermi (www.cref.it) together with several collaborators from various institutions. EFC uses and develops modern data analysis techniques to build economic models based on a scientific methodology inspired by Complex Systems Science (Pietronero 2008) with particular attention to quantitative tests and specific forecasts that provide a solid scientific framework. It consists of a scientific and algorithmic approach that considers specific and concrete problems without economic ideologies and acquires information from the previous growth data of all countries with Artificial intelligence methods, in particular Complex Networks, Algorithms and Machine Learning. Its main features are scientific rigor, precision in analysis and forecasting, transparency and adaptability. The Fitness algorithm overcomes the conceptual and practical problems of the first attempts in this field, and lays the foundations for a testable and successful implementation of the field of Economic Complexity. In particular, it attributes the right importance to the fundamental concepts of diversification and complexity of products. Fitness is in fact defined as the diversification of products weighted by their Complexity (Tacchella *et al.* 2012; Cristelli *et al.* 2013) and the new algorithm allows to obtain both the Fitness of all countries and the Complexity of all products. Diversification characterizes the variety and therefore the stability of the economic system while the complexity of the products describes their technological level and the exclusivity that are connected to the added value and profit. EFC also represents a fascinating and effective bridge between the methods of hard sciences and the problems of the humanities, social sciences and economics. It allows us to tackle some of the most important problems of humanity such as the development and industrialization of countries from a new scientific perspective with concrete results directly useful for progress.

The ideological debate on which is the best theory for economic development is replaced by a new paradigm. There is no ideal theory valid for every situation. As in medicine, where one must first analyze and identify the disease and then decide on the appropriate therapy, there is no universal medicine valid for all diseases. Similarly, there are no universally valid development criteria for the economic development of a country. The various strategies are intrinsically dependent on the context and will vary according to the economic characteristics and circumstances of each country. It is easy to verify from historical data that, for example, the application of neoliberal concepts, i.e. leaving everything on the market without government industrial policy, has been unsuccessful in many cases. But even in those countries where there has been industrial policy, the results have been mostly disappointing. The problem is that they

tried to adapt the same criteria to very different situations.

The EFC methodology allows for a detailed analysis of each country at a given time, defines its industrial competitiveness for each sector and identifies possible realistic development trajectories. These can be different, more or less ambitious, and for each of them it is possible to define the probability of success and the relative increase in production complexity and therefore of the overall Fitness (Tacchella *et al.* 2023). These concepts have interesting similarities with the New Structural Economics by Justin Lin (2014), former Chief Economist of the World Bank in Washington and now at the University of Beijing and also have some affinities with the theories developed by Mariana Mazzucato of London. From the Data Science point of view, the EFC methods are complementary to the NSE economic theory and according to Justin Lin: "Our dream is that this collaboration will lead to sustainable economic development and a world free from poverty".

The EFC methodology allows a radically new approach to the issue of forecasting GDP growth, which represents a benchmark for the quality of the analysis. The review of GDP in the medium to long term is in fact a notoriously difficult problem even for the most competent institutions. The EFC method bases its forecasts on the flow trajectories of countries in the new space defined by GDP (per capita) and Fitness and uses the methods of Dynamical Systems Physics. The result is better than that of the International Monetary Fund (IMF) of Washington (Cristelli *et al.* 2015; Tacchella *et al.* 2018). In fact, after a detailed analysis, Bloomberg concluded that: "New research has demonstrated that the 'fitness' technique systematically outperforms standard methods, despite requiring much less data". Note the emphasis on the fact that few data are used.

In fact, the IMF has about 2,000 people dealing with this problem and, in general, there is a team for each country that uses country-specific analysis criteria. Then there is the core team in Washington that tries to harmonize all these analyses. Clearly, all available data is used in these analyses but, as we will see below, this can lead to advantages and disadvantages. Furthermore, such a method is inherently characterized by subjective elements and is therefore difficult to reproduce or test in a scientifically independent manner. Another element of the EFC method is that it can correctly explain and predict the fantastic development of China over the past forty years (Pietronero *et al.* 2025). A fact that has eluded most experts and traditional methods. Finally, Fitness appears to be a much more appropriate criterion of validity for the development of countries than GDP and has been adopted for the analysis of IFC-World Bank projects, EU NRRP projects and London EBRD projects.

Big Data, opportunity and myth

The immense accumulation of data is a new phenomenon that induces many considerations, represents great potential and also leads to mythical expectations. A few years ago, it was even suggested that if you have enough data, you can do without theories and traditional understanding. This extreme view now seems to have waned and has been replaced by Large Language Models. The basic idea in these models is to catalogue an immense number of texts on a neural network that provide the basis of their knowledge. At first, the criterion for training was extremely simple and consisted in guessing the missing word in a text. With the Transformer, a fundamental step forward has been taken and, in addition to reproducing and summarizing texts admirably, LLMs have begun to answer questions in an almost unexpected way. Recently, this trend seems to be moving even towards a sort of 'reasoning'. Many of these features were not anticipated by the developers themselves and were quite surprising. It therefore seems that when the system reaches an enormous complexity, emergent behaviors are spontaneously activated. A fascinating problem from the scientific side in which the Science of Complexity seems to be the appropriate context for their understanding.

This approach is typical of Big Science, since the amount of data needed and the related energy and costs can be astronomical, even in the order of billions of euros.

However, we have also seen complementary situations, such as the Google algorithm, developed by two students in a garage and the Fitness algorithm in which even the data must be carefully limited and selected. However, this method gives superior results to the main economic institutions that certainly use all available data. How is this apparent

paradox possible? First, it should be considered that we are now trying to answer a very precise scientific question, albeit a far-reaching one. This answer is clearly not contained in the data available even in the most powerful LLMs as we are really trying to increase our knowledge.

The standard analysis of countries 'competitiveness considers a series of elements such as education, finance, transport, production, export, pollution etc. and with an adequate weighting of these elements leads to an overall score for the country. In the end, this analysis requires defining over 100 parameters to assign the weight relative to all these elements, each of which is in turn characterized by a large amount of data. Clearly, this is a task that introduces subjective and ideological elements and that cannot adequately consider all the possible interactions between these elements. Furthermore, the analysis is carried out for each country individually and only at the end are different countries compared.

The Economic Complexity (EFC) (Pietronero *et al.* 2024) method corresponds to a change of perspective that goes beyond individual analysis. All countries are considered as nodes of an integrated network, and the connections are given by the products they produce. In practice, we consider the bipartite network of countries and products, and the idea is to extract interesting information from the general properties of this network in the spirit of the Google algorithm, although, as we shall see, a different algorithm is needed here. The first problem is that consistent and homogeneous data are needed for all countries, but they are usually not available or are only partially available. So, the new vision immediately shows the limits of the available data. It is a qualitative problem, not dependent on the amount of data. So, we start working with what is available, but we would like to have something else. For example, it would be very useful to have complete information for all countries of the work activities, but this data is not currently complete, so we use the data of the exported products for which we have complete and homogeneous information as a proxy for competitiveness.

In principle, clearly you have more information than simple products, but the problem is that these data are not independent and considering them all together only leads to confusion or an intrinsically subjective interpretation, as it requires assigning parameters and weights to the various data. But we wanted to go beyond this phase and do a scientific analysis, that is, provided a unique, reproducible result that was not dependent on subjective interpretation. This leads to a selection of the data and to a drastic reduction of the same. Only a selected subset is really useful (Lin 2014; Cristelli *et al.* 2015), adding more data would lead to a sort of confusion.

This discussion is analogous to Poincaré's studies on predictability in dynamic systems, which is high if the dimensionality of these systems is low. In some ways, having data in many directions is similar to having a high dimensionality. Therefore, increasing the data increases the information but also the mathematical noise or, in a more colloquial way, the confusion. In the case of Fitness, it has been shown that the optimal situation is with data appropriately selected with respect to the homogeneity criteria. Here we are clearly in the field of Small Science, which requires a specific culture to be developed in the field of Complex Systems Science.

It is easy to imagine that many of the open scientific problems in various fields where a lot of data is available, in particular, for example, in medicine, can have considerable advantages from algorithmic approaches similar to Economic Fitness. As we have seen in situations of this kind, a specific and targeted approach to the problem to be considered is necessary. This is a qualitatively different problem from the use of LLM models as the goal is to obtain new information that is not contained in the dataset with which the network was trained. These types of developments have the characteristics of Small Science and would seem particularly suitable for Italy as they depend on ingenuity and creativity, typical of Small Science, rather than on brute force or material resources.

4. The impact of AI on the world of work

This is a typical case where analysis and forecasts are enormously divergent. Some think that human work will be almost entirely replaced by AI and robots and envision apocalyptic scenarios with a huge number of unemployed people and a few owners of these technologies that will upset and dominate the world, while others see in AI a great help and support to many activities that will become easier without eliminating employment.

Let us try to learn some lessons from some specific cases of innovations that have produced

disruptive developments. In 1997, Kodak had about 160,000 employees. And about 85% of the world's photography was taken with Kodak cameras. With the rise of digital cameras in recent years and now even with photos from cell phones, Kodak has completely failed, and all its employees have been laid off.

Many other more recent examples can be made, such as Nokia, which at one point was a leader in mobile phones and has now disappeared. The curious thing about Kodak, and perhaps not well known, is that Kodak itself had a research group that contributed to the development of digital cameras but, since the dominant industrial culture was that of chemistry, the digital line did not have much following.

In the early 1980s, IBM developed the first Personal Computers, but the dominant culture was that of large mainframe computers and the PC was considered little more than a toy. It is from this perspective that the operating system of IBM PCs was contracted and practically given to Bill Gates who made his fortune.

Finally, in the 1970s, Xerox developed the graphical interface without realizing that this would lead to the Mouse with all its related developments. This technology was then given to Steve Jobs with the consequences we all know. Even Xerox is now a pale memory of the great innovative company it was many years ago.

These examples show that a certain culture and business tradition can represent a brake on innovation and that innovative companies must be ready to question themselves continuously to seize the technological opportunities that they themselves have sometimes developed. The situation of Nokia is quite different because, if on the one hand it was impossible to predict that a group of brilliant Finnish engineers would develop a particularly efficient mobile phone, on the other hand it was possible to predict, with the methods of Economic Fitness, that the evolution from the simple mobile phone to today's smartphones would have been problematic for Nokia. In fact, this step implies that the phone is integrated with many other technologies, and these were certainly present in the industrial ecosystem of Silicon Valley but not in Finland.

Returning to the question of the impact of AI on the world of work, some predictions are reasonably realistic while others are very difficult. For example, it is very difficult to predict which new activities can be developed and invented, because the potential of these new methods is in tumultuous development and therefore the best thing is to stay informed about these developments, and their possible consequences.

The problem of the impact of AI on existing work activities is different and much more manageable. Also, for this case, opinions and expectations are quite different. However, here we can develop an original contribution that uses both the Fitness methodology and the LLM models in an interesting combination (Fenoaltea *et al.* 2025).

Each work activity is associated with a certain number of skills or abilities that are necessary to carry it out. All work activities and related skills are codified by international organizations and the data are public. Looking at these data, we can observe that some activities require many skills, then progressively others require fewer skills, up to some that use very few. The mathematical nature of this data distribution is *nested*, quite similar to that between countries and products that stimulated the development of the Fitness algorithm. Using this algorithm we then obtain the Fitness of all work activities and the Complexity of all skills. For example, one of the activities with the highest Fitness is that of an airplane pilot, as it requires many skills, some of considerable complexity. An opposite example of activities with low Fitness is that of the model, which requires very few skills.

Then there are data that estimate the impact that AI will have on each skill. These data are of a subjective nature, obtained by interviewing experts. By combining all this, we can then obtain the estimated impact of AI on each work activity. The general trend is that the greater the Fitness, the lower the impact of AI. This trend is understandable because, if you have many skills and perhaps even complex ones, it is difficult for AI to have a great impact on all these skills.

But let's also observe some peculiarities and notable outliers. For example, some activities that would be said to be of a certain complexity such as those of physicists, mathematicians and computer scientist have a low Fitness and therefore a high impact expected from AI. This is due to the fact that the skills classified at the moment correspond mostly to manual and practical activities, while conceptual ones are relatively few. This anomaly can be cured by increasing the skills of intellectual

activities in order to balance the distribution and we are working on this improvement.

Regarding the outliers, the airplane pilot who has the highest Fitness is quite replaceable with AI. This is due to the fact that, although it is complex to replace a pilot, there is a great effort and desire to do so using drones.

An opposite case is that of judges, who in theory are easily replaceable using an LLM model that learns all the laws, compares them with the crimes committed and produces the sentences. However, the idea of replacing a judge with AI does not appear socially acceptable. This is why AI will not have much impact on these activities, despite the theoretical possibility of doing so.

We have therefore seen that considering only the theoretical and subjective impact of AI on the various skills is not very realistic. It is much more appropriate to consider the investment that is actually made for these substitutions. We therefore considered a series of start-up companies, and, through LLM methods, we compared their activities with the definition of the various skills. In this way, we obtain a criterion of substitutability or impact of AI based on the investments that are actually made rather than on theoretical and subjective opinions. In this way, we obtained an index of the impact of AI on skills, and therefore on work activities, which at the moment presumably represents the state of the art in the field (Fenoaltea *et al.* 2025).

5. Comparison between LLM analyses and the creation of new information

We have seen that LLM models are based on neural networks that emulate the architecture of the brain, are of a generalist nature, and require enormous amounts of data and energy. In 1996, Google's Page Rank algorithm made it possible to define a hierarchy of all Internet sites connected with the question that arises. Now with LLM, we can enter the content of these sites and acquire all the information (text) they contain, which is then appropriately stored. In principle, all the public information in the world will soon be stored. With Transformer and subsequent developments, a further step is taken that allows for high-quality reasoned summaries. What is surprising, however, is that these systems seem to go even further and become capable of answering questions, which were somehow not foreseen in the original programs. Furthermore, it is observed that the ability to answer some specific questions, for example in mathematics, becomes possible in a discontinuous way only after the acquisition of billions of data. These phenomena appear as emerging and indicative of a certain '*reasoning*' ability of these systems, which were initially trained only by the criterion of adding the missing correct word to a text. At the moment, it is difficult to have a clear idea of where we can get through these developments and in what sense we can talk about the emergence of elements of creativity.

On the other hand, we have seen a qualitatively different example with Economic Fitness. In this case, we are faced with a very precise and well-defined problem, albeit far-reaching, which is how to plan and predict the economic development of countries with data-driven scientific methods. Curiously, here it is important to limit the data to the most significant ones and invent a new algorithm. Then a new space is introduced to analyze the dynamics of countries and from this, with methods inspired by Physics, it is possible to make analyses and forecasts that are superior to those of standard institutions that use much more data. In this case, however, we are using existing data as a basis for acquiring new knowledge with a new method that cannot be found in the existing literature, so it is natural that LLM models cannot reproduce these results that correspond to the acquisition of new knowledge.

However, considering the elements of creativity that we discussed earlier, it is legitimate to ask whether in the future LLM models will be able to autonomously develop concepts and methods that do not currently exist. Innovation often consists in combining different technologies that had not been combined before, such as in autonomous driving. It is reasonable to think that this type of innovation may one day be within the reach of LLM models. For more radical innovations such as Economic Fitness compared to standard economic models, the situation appears more difficult, but it is unwise to be too drastic with these arguments because this field has already accustomed us to surprising and unexpected developments.

At the moment, we can conclude that the development of general methods of the LLM type seems associated with large investments and large energy capacities, although the Chinese company Deepseek seems to have brought significant improvements in efficiency.

In a complementary way, the exploration of scientific problems based on Data Science seems to require a specific creative effort whose results are not reproducible by LLMs. For example, if we compare our results of the impact of AI on the labor market reported in Fenoaltea *et al.* 2025 with the same question asked to ChatGPT, it is easy to realize that our new results are far superior to what can be achieved with ChatGPT at the moment.

In conclusion, the development of generalist LLM methods seems to belong more to Big Science, while the development of new scientific knowledge, beyond the current one, requires ingenuity and creativity that seem closer to the culture of Small Science.

6. What Italy has already done in AI

Italy has developed an AI ecosystem with discrete contributions in both research and industry. National academic research is of a high standard: in 2022, the country produced 3,261[1] scientific publications on AI, ranking seventh in the world. Several centers of excellence participate in European AI projects (Italy is involved in 12% of EU projects). On the private side, the Italian AI market is growing rapidly, reaching about 1.2 billion euros[2] in 2024 (+58% in one year).

However, the entrepreneurial fabric remains fragmented: there are only about 600 patents by Italian inventors and just over 350 AI startups founded since 2017, a figure that places Italy at the bottom of the list in Europe. The government has launched a National AI Strategy and targeted initiatives (e.g., the FAIR foundation with 28.7 million from the PNRR) to stimulate innovation. Critical issues remain: the difficulty in retaining specialized talent and the absence of global industrial champions in the sector. At the international level, Italy's strengths and weaknesses compared to AI leaders are highlighted. The United States and China dominate thanks to massive investments and technology giants: in 2022, the U.S. registered over 16,000 AI patents, compared to less than 800[3] in the entire European Union. But China seems to be doing even better. Considering the patents approved in the field of AI, 61% are Chinese, 21% US and only 2% EU and UK. The UK and France have ambitious strategies and centers of excellence: the former was the cradle of DeepMind, the latter announced a 109 billion euros plan for AI. Italy can count on quality scientific research and participation in European programs, but it suffers from the lack of global industrial champions and advanced computing infrastructures (the lack of data centers and supercomputers is a significant obstacle). The Italian AI landscape is lively, but it needs to accelerate to bridge the gap with leading countries by strengthening investments and skills.

7. How to make Italy a leader in AI

We have discussed the origin and the tumultuous development of AI in recent years that show unexpected and almost mysterious emerging properties. The field is fascinating both from a purely scientific point of view and for the impact it will have on work activities and on the whole of society.

From a scientific point of view, it is important to create an interdisciplinary culture that allows an original approach, appropriate to this field. Certainly, the Physics of Complex Systems appears as an appropriate context to try to define a cultural and scientific framework for the many features of the AI field. From the methodological standpoint, we have discussed situations that require huge amounts of data and energy but also those in which fundamental contributions can be made by two students in a garage, like in the case of Google, but also for the economic Fitness algorithm in which the drastic reduction of data was actually essential. Clearly there is room for both these extreme situations and also for all those in between.

This situation recalls the dichotomy between Big Science and Small Science in Physics. On the one hand, Big Science can be planned and involves the development of large laboratories that have great media and political visibility. On the other hand, Small Science often concerns situations that

1 Strategia italiana per l'intelligenza artificiale: facilitatori per le imprese e sostegno alle startup, Manuela Perrone, in *Il Sole 24 Ore*, 22 luglio 2024, https://www.ilsole24ore.com/art/intelligenza-artificiale-facilitatori-le-imprese-e-sostegno-startup-AFl7EyzC.

2 Intelligenza artificiale, boom del mercato italiano: +58%, 1,2 miliardi di euro, *Polimi School of Management*, February 2025, https://www.osservatori.net/comunicato/artificial-intelligence/intelligenza-artificiale-italia/.

3 Il ritardo dell'Europa nell'intelligenza artificiale: numeri e sfide, Biagio Simonetta in Il Sole 24 Ore, 23 agosto 2024, https://www.ilsole24ore.com/art/la-sfida-sull-intelligenza-artificiale-ecco-numeri-ritardo-europeo-AF6470UD.

are not very predictable and plannable and that individually are less spectacular. But in science one wants to explore an unknown area and therefore evaluating it carefully with known criteria can be very problematic. In fact, many of the discoveries in the field of Small Science are unexpected and difficult to evaluate a priori. It should also be noted, though, that Nobel prizes in Physics are much more numerous in Small Science than in Big Science. Also, in the context of AI, these apparently opposite trends are well identifiable and are actually complementary. Both should be cultivated with care and efficiency without deluding ourselves that size and power are the only keys to success. In particular, Italy can hardly compete in the context of AI Big Science, while the development of a Small Science culture for AI, which has excellent prospects for being competitive at the highest level, appears much more realistic and appropriate.

From an educational standpoint, it is important to train a generation of young scientists who can make an original and creative contribution to the field of AI combined in various areas and disciplines in an interdisciplinary context. The development of LLM techniques and those that will follow will involve a limited number of specialists. But the use of these techniques in all fields, both industrial and scientific, will involve a much larger number of people. A broad cultural investment must therefore be made in this direction.

Below is a preliminary analysis but, in case of interest, it would be very easy to produce a 3-month study with a larger research group.

Economic Complexity teaches us that the higher the Fitness of an ecosystem, the higher its growth potential. Fitness is high when an ecosystem has internally everything that is necessary for success without having to purchase goods and knowledge from the outside because this would lead to a slower speed in adopting innovation.

Artificial intelligence requires the following resources and skills:

Resource / Expertise	Prevailing Investment	EU	USA	China
Raw materials for chip construction	Infrastructure	Poor	Good*	Excellent
Chip design	Human capital	Poor	Excellent	Good
Lithography for semiconductors <=2nm	Infrastructure	Excellent	Poor*	Poor*
Chip manufacturing <=2nm	Infrastructure	Poor	Poor*	Excellent
Chip assembly	Infrastructure	Poor	Poor*	Excellent
Hardware acceleration software	Human capital	Poor	Excellent	Poor*
Datacenters for training (cold climate and basic energy)	Infrastructure	Poor	Good	Good
Global datacenters for inference (latency, base energy)	Infrastructure	Poor	Excellent	Good**
Basic research on AI	Human capital	Good	Excellent	Excellent
Risk Capital (All stages)	Human capital	Poor	Excellent	Good
Creation of Foundational Models and access to data	Infrastructure	Poor***	Excellent	Excellent
Creation of Vertical Models	Human capital	Good	Good	Good
Creating AI Applications	Human capital	Good	Excellent	Excellent
Integration and Platforms	Human capital	Good	Excellent	Excellent
Go-To-Market and Company Scale-ups	Human capital	Poor	Excellent	Good
Protection of Dominant Positions (M&A, Investments, Lobbying)	Human capital	Poor	Excellent	Good

* Indicates that the rating could improve rapidly thanks to investments already made. The US, with the CHIP Act, is investing in chips and alternatives to EUV lithography. They have also entered into direct agreements with Intel, TSMC, Samsung, TI and others.

** China was rated good and not excellent because the US does not allow Chinese companies to enter the US market easily.

*** In addition to infrastructure investments, profound regulatory changes are needed. Mistral, the only foundational LLM player, has investments orders of magnitude smaller than its US and China competitors.

For human capital it is good to note that Italy potentially has good access but the flight of junior and qualified talent (550,000 talents lost in the last three years according to the CNEL study) makes the country absolutely devoid of the necessary resources. Furthermore, lacking the qualified ecosystem, the talent that remains grows in skills at a lower speed and totally lacks the more senior and managerial figures who today have only been trained abroad.

It seems clear that limiting the brain drain, aggressively repatriating senior talent and becoming more attractive than the US and China for global talent is the only way to have an opportunity.

Attracting talent is given by several elements but the most important are:

- *Values*. Prove to be a country that appreciates young talent and creates more favorable and even less restrictive research and application conditions between the US and China. Today the critical issue is that Italy has always sent the opposite signal, making regulations against digital and innovation, mainly protecting the status quo. And this requires perhaps the most profound and difficult change to achieve.
- *Infrastructure*. Make the infrastructure investments necessary to increase fitness.
- *Public Research.* Create a job offer that is economically and socially more convenient than the American and Chinese ones. Here the critical issue is that AI talent today is worth and is also paid an order of magnitude more than the average. The critical issue is to bring together researchers who earn 25,000 euros a year with those attracted who will necessarily earn 250,000 euros a year.
- *Private Human Capital.* Provide incentives to businesses to bring back talent. The critical issue will be that the private market will want to invest only when the market is clear, thus creating a stalemate.

References

Cristelli M., Gabrielli A., Tacchella A., Caldarelli G., Pietronero L. (2013), Measuring the intangibles: a metrics for the economic complexity of countries and products, *PLoS ONE*, 8, n.8, e70726 <https://doi.org/10.1371/journal.pone.0070726>

Cristelli M., Tacchella A., Pietronero L. (2015), The Heterogeneous Dynamics of Economic Complexity, *PLoS ONE*, 10, n.2 <https://doi.org/10.1371/journal.pone.0117174>

Fenoaltea E.M., Mazzilli D., Patelli A., Sbardella A., Tacchella A., Zaccaria A., Trombetti M., Pietronero L. (2025), Follow the money: a startup-based measure of AI exposure across occupations, industries and regions, *Arxiv preprint* <https://doi.org/10.48550/arXiv.2412.04924>

Lin J. (2014), Industrial policy revisited: a new structural economical perspective, *China Economics Perspective*, 7, n.3, pp.382-396

Pietronero L. (2008), Complexity Ideas from Condensed Matter and Statistical Physics, *Europhysics news*, 39, n.6, pp.26-29

Pietronero L., Tacchella A., Mazzilli D., Zaccaria A., Pugliese E., Cader M., Lin J.Y. (2025), The fantastic growth of China and the middle income trap in the light of economic fitness and complexity, *preprint*

Pietronero L., Tacchella A., Zaccaria A., Gabrielli A., De Marzo G., Mazzilli D., Patelli A., Saracco F., Sbardella A., Benzi R., Cimini G., Pugliese E., Cader M. (2024), *Economic Fitness: Concepts, Methods and Applications* <http://www.lucianopietronero.it/wp-content/uploads/2024/08/Economic-Fitness-and-Complexity1.0_2024.pdf>

Tacchella A., Cristelli M., Caldarelli G., Gabrielli A., Pietronero L. (2012), A new metrics for countries'fitness and products'complexity, *Scientific Reports*, n.2, 723 <https://doi.org/10.1038/srep00723>

Tacchella A., Mazzilli D., Pietronero L. (2018), A dynamical systems approach to gross domestic product forecasting, *Nature Physics*, 14, pp.861-865

Tacchella A., Zaccaria A., Miccheli M., Pietronero L. (2023), Relatedness in the era of machine learning, *Chaos, Solitons and Fractals*, 176, 114071

The Nobel Committee for Physics (2021), *Scientific Background on the Nobel Prize in Physics 2021 to G. Parisi "For groundbreaking contributions to our understanding of Complex Physical Systems", The Royal Swedish Academy of Science, Stockholm, Sweden* <https://www.nobelprize.org/uploads/2021/10/sciback_fy_en_21.pdf>

The Nobel Committee for Physics (2024), *Scientific Background on the Nobel Prize in Physics 2024 to John J. Hopfield and Geoffrey Hinton "For foundational discoveries and inventions that enable Machine Learning with Artificial Neural Networks", The Royal Swedish Academy of Science, Stockholm, Sweden* <https://www.nobelprize.org/uploads/2024/11/advanced-physicsprize2024-3.pdf>

Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser Ł., Polosukhin I. (2017), Attention is all you need, *arXiv preprint* <https://doi.org/10.48550/arXiv.1706.03762>

Luciano Pietronero
luciano.pietronero@cref.it
Emeritus and member of the Scientific Council of the Enrico Fermi Research Center (CREF), where he served as President from 2019 to 2024. He has contributed significantly to both academic and industrial research, promoting interdisciplinary approaches and founding the CNR Institute for Complex Systems in 2004. Professor of Physics at Sapienza University (1986 -2020), he fostered collaboration with CNEL to apply economic fitness methods for more rigorous economic analysis. He was recently appointed Professor at the Universities of Taiyuan and Hangzhou (China), within the International Research Centre for Complexity Sciences.

Marco Trombetti
marco@picampus.it
Entrepreneur in the technology sector, founder and Chief Executive Officer of Translated, an online translation platform recognized as one of the most influential translation companies worldwide. In 2007, he founded Memopal, where he currently serves as President and CEO. Memopal develops software that automatically backs up users' files to the cloud. Under his leadership, Memopal has fostered a work environment that has been ranked among the 30 best workplaces in Italy.

L'Intelligenza artificiale generativa sta ridefinendo una nuova era per la ricerca accademica?

Filippo Giordano
LUMSA

Saverio Lovergine
INAPP

Per indagare l'impatto di strumenti di Intelligenza artificiale generativa sull'integrità della ricerca, si analizzano le linee guida disponibili sui siti web di 20 delle migliori università europee. L'analisi ha evidenziato che le università spingono all'utilizzo di tale tecnologia, nelle more di un uso appropriato e trasparente in linea con i principi etici e di integrità della ricerca. Nelle conclusioni si propongono sia linee guida dinamiche, adattive e aggiornabili, in risposta ai rapidi progressi di tali tecnologie, sia standard condivisi dagli stakeholder dell'ecosistema della ricerca che garantiscano un utilizzo responsabile ed equo di tali tecnologie.

This study investigates the impact of generative Artificial intelligence tools on research integrity by analysing the guidelines published on the websites of 20 leading European universities. The analysis reveals that universities encourage the use of such technology, provided it is applied appropriately and transparently, in line with ethical principles and standards of research integrity. The conclusions proposes both dynamic, adaptive, and updatable guidelines capable of responding to the rapid evolution of these technologies and shared standards among research ecosystem stakeholders to ensure responsible and equitable use.

DOI: 10.53223/Sinappsi_2025-02-3

Citazione
Giordano F., Lovergine S. (2025), L'Intelligenza artificiale generativa sta ridefinendo una nuova era per la ricerca accademica?, Sinappsi, XV, n.2, pp.31-45

Parole chiave
Etica
Intelligenza artificiale
Ricerca universitaria

Keywords
Ethics
Artificial intelligence
University research

Introduzione

Il rapido sviluppo di strumenti di Intelligenza artificiale generativa (GenAI) sta definendo scenari inediti, suscitando sensazioni e atteggiamenti contrastanti nel dibattito pubblico: da un lato, l'entusiasmo per le potenzialità trasformative di tali tecnologie; dall'altro, le preoccupazioni per le implicazioni etiche, epistemologiche e di integrità derivanti dal loro utilizzo. Un esempio emblematico per comprendere le sfide poste dall'adozione di strumenti GenAI è rappresentato dal recente caso (gennaio 2025) dell'Università di Ferrara, dove un'intera sessione d'esame è stata annullata dopo che si è scoperto che una parte degli studenti aveva utilizzato ChatGPT per rispondere alle domande. Oramai viviamo in una *exponential growth of technology era* (Aspen Institute Italia 2024), un'epoca in cui l'evoluzione tecnologica alimenta la diffusione di sistemi sempre più sofisticati, in grado di interagire con gli esseri umani e di supportare attività complesse attraverso modalità interattive, adattive e scalabili (Amir *et al.* 2016).

Sebbene il concetto di Intelligenza artificiale (IA) sia stato introdotto da diversi decenni, è solo in tempi

recenti che questa tecnologia ha conosciuto una larga diffusione, affermandosi come un elemento cruciale che sta rapidamente cambiando il modo in cui le transazioni e le interazioni sociali, culturali ed economiche sono organizzate nella società odierna.

La capacità di rivaleggiare, se non di imitare, l'intelligenza umana nella risoluzione di problemi complessi distingue l'IA dalle altre tecnologie, poiché consente alle macchine di svolgere molti compiti cognitivi, tradizionalmente riservati agli esseri umani (Osoba e Welser 2017; Bathaee 2018; Sætra 2020). L'avvento di ChatGPT nel 2022 ha rappresentato un punto di svolta nella storia di tale tecnologia, evidenziando le straordinarie capacità di GenAI nel supportare processi decisionali e nel generare contenuti testuali, visivi, sonori e multimediali originali con un livello di complessità paragonabile a quelli originati da esseri umani. Gli strumenti GenAI identificano e riproducono schemi, modelli, stili e strutture presenti nei dati di addestramento (Baidoo-Anu *e* Owusu 2023; Halaweh 2023; Kaplan-Rakowski *et al.* 2023). Oltre a ChatGPT, altre piattaforme GenAI – come Perplexity, Copilot, Claude, DeepSeek ecc. – ridefiniscono i confini tra ciò che è artificiale e ciò che è reale, trasformando il modo in cui apprendiamo, insegniamo e facciamo ricerca scientifica.

Alla base degli strumenti GenAI risiedono i cd. *foundation models*, modelli *general-purpose* addestrati su dati di grandi dimensioni con tecniche di autoapprendimento, progettati (e perfezionati) per svolgere un'ampia gamma di compiti anche non previsti in fase di addestramento (Brown *et al.* 2020). Nonostante la loro ampia diffusione, non è ancora ben chiaro come questi modelli funzionano, falliscono e di cosa siano capaci di fare grazie alle loro proprietà emergenti (CRFM 2022). La distinzione concettuale tra *foudation models* e le piattaforme GenAI è di natura funzionale: i primi costituiscono l'infrastruttura modellistica sottostante, mentre le seconde definiscono una specifica modalità di impiego volta alla sintesi creativa dell'informazione.

Gli strumenti GenAI non costituiscono entità neutre; infatti, le loro prestazioni possono dipendere dalla qualità, varietà e distribuzione dei dati utilizzati in fase di training, nonché dai metodi di ottimizzazione e di tuning, che possono riflettere i pregiudizi dei loro sviluppatori (von Krogh *et al.* 2023). Analogamente ai *foundation models*, anche gli strumenti GenAI presentano prestazioni in termini di accuratezza e affidabilità spesso imprevedibili (Chan 2023), con la possibilità di generare contenuti plausibili ma falsi, errati o privi di senso. Questo fenomeno, noto come allucinazioni dell'IA (*AI hallucination*) è stato ampiamente documentato (Aljamaan *et al.* 2024). Inoltre, tali modelli hanno sollevato criticità in termini di opacità algoritmica, poiché i processi decisionali interni risultano non osservabili né spiegabili, trasformando l'algoritmo in una vera e propria scatola nera (cd. problematica del black box), soprattutto per gli utenti privi di competenze o formazione tecnica specifica in materia.

In virtù di ciò, l'Unione europea è intervenuta con l'approccio normativo pioneristico dell'AI Act[1], che definisce un quadro legislativo per gestire i rischi e promuovere un uso responsabile e sicuro di tale tecnologia, nonché garantire la sicurezza e il rispetto dei diritti fondamentali di cittadini e imprese.

In ambito accademico e della ricerca, il dibattito in letteratura si concentra sull'utilizzo di GenAI nelle diverse fasi del processo scientifico, evidenziando come tale impiego possa variare in funzione della disciplina di riferimento. Anche se i confini tra le diverse fasi della ricerca sono spesso sfumati, tale articolazione consente di distinguere chiaramente le potenzialità, i compromessi e le responsabilità associate all'adozione di GenAI nel contesto accademico (Cornell University Task Force 2023).

In questa direzione, la letteratura si è anche concentrata sui potenziali impatti sulla creazione di conoscenza, sul lavoro accademico e sulla credibilità e integrità dell'ecosistema della ricerca (Antonenko e Abramowitz 2023).

Per quanto riguarda la creazione di conoscenza, Floridi (2025) introduce il concetto di *distance writing* (o *wrAIting*), quale pratica letteraria emergente in cui l'autore umano non funge da produttore testuale diretto, ma da architetto o progettista delle possibilità narrative (definisce requisiti, vincoli e orientamenti

1 Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i Regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (Regolamento sull'intelligenza artificiale), GU L, 2024/1689, 12.7.2024.

stilistici) del contenuto generato dagli LLM. Questo cambio di paradigma rappresenta non solo un intervento tecnologico nel processo di scrittura, ma una concettualizzazione fondamentalmente diversa di cosa significhi scrivere un testo. Basandosi su un esperimento denominato *Encouters*, lo stesso Floridi sostiene che la *distance writing* rappresenta un'evoluzione significativa nell'autorialità in quanto essa non sostituisce, ma espande la creatività umana all'interno di un quadro progettuale sollevando, allo stesso tempo, domande essenziali sui confini e sul futuro della letteratura umana in una cultura guidata dall'IA.

In virtù di ciò, Hutson (2024) evidenzia come l'attuale impianto normativo risulti inadeguato a seguito dell'introduzione di GenAI che, a differenza dei precedenti strumenti digitali, crea contenuti originali attraverso complessi processi algoritmici, confondendo i confini tra paternità e originalità. Attualmente la normativa in materia di proprietà intellettuale si scontra con la realtà emergente di co-creazione tra esseri umani e GenAI. In tale contesto l'autore propone una riforma del diritto d'autore orientata verso un modello adattivo, capace di integrare l'IA come strumento creativo complementare, ma non sostitutivo dell'ingegno umano. Ciò implica l'adozione di provvedimenti giuridici che valorizzino l'intervento umano nella fase ideativa, nella generazione degli input o nella post-produzione, come condizione necessaria per l'attribuzione della titolarità della produzione.

A ciò si aggiungono ulteriori rischi derivanti dall'uso di GenAI quali, ad esempio, il plagio (tanto che qualcuno afferma che viviamo in un'epoca di post-plagio (*postplagiarism*[2])) e la gestione impropria di informazioni sensibili (Ghimire e Edwards 2024).

Su queste problematiche, e altre di natura etica, metodologica ed epistemologica legate all'autenticità e all'integrità dei risultati prodotti da strumenti GenAI, molte istituzioni universitarie, della ricerca ed educative hanno risposto in modo difforme all'introduzione di tale tecnologia: vietando completamente l'uso di tali strumenti; ignorando tale problematica; introducendo politiche di condotta e Linee guida (LG) per disciplinarne l'utilizzo, al fine di garantire trasparenza e integrità nella produzione scientifica. A tali reazioni si aggiunge anche quella della rilevazione, in modo affidabile, dei contenuti generati da GenAI. Diverse aziende hanno tentato di sviluppare strumenti di identificazione e riconoscimento di tali contenuti, con risultati poco soddisfacenti. OpenAI nel 2023 ha interrotto il proprio progetto di riconoscimento di contenuti prodotti da ChatGPT a causa dell'elevato numero di falsi positivi e negativi riscontrati nell'analisi dei prodotti. Per problemi analoghi anche Turnitin[3] ha sospeso l'uso del proprio strumento. *Originality.ai*, pur avendo mostrato una maggiore efficacia e tassi di successo, ne ha sconsigliato l'utilizzo in ambito accademico[4]. Jiang *et al.* (2024) hanno riportato l'esperienza di aziende che hanno utilizzato la tecnica della filigrana (*watermark*), già utilizzata per trarre informazione sull'origine e la provenienza delle informazioni di file multimediali o di altro genere, per identificare contenuti generati dall'IA. La ricerca ha evidenziato che tramite tale tecnica l'attribuzione è altamente accurata quando il contenuto generato dall'IA non è sottoposto a post-elaborazioni. Pertanto, anche per questa tecnica si evidenzia la necessità di sviluppare filigrane più robuste per garantire un'attribuzione affidabile. L'importanza dell'attribuzione trova ulteriore conferma nel fatto che spesso i modelli GenAI imputano erroneamente a sé stessi testi originali, inclusi estratti di opere famose. L'inaffidabilità di tali strumenti di attribuzione ha evidenziato le criticità del loro utilizzo; uno degli esempi più eclatanti in letteratura è rappresentato dal caso dell'istruttore della Texas A&M University, Dr. Jared Mumm, che ha bocciato gli studenti di un intero corso basandosi erroneamente sull'utilizzo di ChatGPT per verificare i saggi prodotti dai propri studenti (Agomuoh 2023).

Sulla base di queste considerazioni, emerge la necessità di adottare politiche e LG chiare e condivise per un uso responsabile di strumenti GenAI nella ricerca e nella scrittura accademica, che diano priorità all'integrità della ricerca;

2 In un mondo post-plagio, i sistemi di IA sono comunemente utilizzati nella scrittura e in altri processi creativi, rendendo difficile distinguere tra contenuti generati dall'uomo e dall'IA. Questa epoca richiede un ripensamento dell'integrità accademica, sottolineando l'uso etico della tecnologia e promuovendo un mondo socialmente giusto (Eaton 2023).

3 Si veda: *Turnitin announces AI writing detector and AI writing resource center for educators*, 13 febbraio 2023.

4 Si veda Gillham J., *Do AI Detectors Work? Open AI Says No - Prove It for Charity*, 2024, https://originality.ai/blog/do-ai-detectors-work.

accelerino l'innovazione; affrontino problematiche quali la privacy e la sicurezza dei dati; riflettano sull'evoluzione positiva e negativa sull'uso di tali strumenti sulle pratiche e sulle competenze delle comunità di ricerca (Cornell University Task Force 2023).

Alla luce di ciò, il presente lavoro intende analizzare le principali LG sull'uso di GenAI emanate in ambito accademico europeo. Dopo la presente introduzione, il paragrafo 1 offre una rassegna della letteratura sulle principali LG redatte da organizzazioni internazionali e istituzioni accademiche; nonché alcuni studi sulle analisi delle informazioni e delle istruzioni presenti nelle LG di prestigiose università sull'uso di strumenti GenAI al loro interno. Il paragrafo 2 descrive la metodologia utilizzata per analizzare, grazie ad una checklist di 18 criteri basata sulla letteratura, le LG disponibili al pubblico sui siti web delle università europee presenti nei tre ranking delle migliori università del mondo. All'esposizione dei risultati (paragrafo 3), seguono una proposizione delle principali evidenze riscontrate (paragrafo 4) e le conclusioni.

1. Revisione della letteratura

Per affrontare la questione dell'avvento degli strumenti GenAI nel mondo accademico è indispensabile una collaborazione multidisciplinare commisurata alla natura fondamentalmente socio-tecnica (CRFM 2022). Pertanto, l'ecosistema della ricerca dovrebbe cooperare per definire politiche di condotta e LG per disciplinare l'utilizzo di strumenti GenAI, al fine di garantire un uso responsabile nella ricerca e nella scrittura accademica.

Ad esempio, gli editori scientifici, tra cui Elsevier, Springer Nature e Sage, hanno già pubblicato specifiche LG rivolte ad autori, revisori ed editori (Nature 2023; COPE 2023).

Anche le organizzazioni internazionali e le istituzioni accademiche hanno contribuito alla definizione di politiche e LG per garantire un utilizzo etico e responsabile di GenAI nell'ecosistema della ricerca scientifica.

L'Unesco ha fornito un quadro di riferimento con il documento *Guidance for generative AI in education and research* (Miao e Holmes 2023), basandosi sulle proprie Raccomandazioni del 2021 sull'etica dell'IA. Le linee guida sottolineano l'importanza di un approccio umano-centrico all'IA che tuteli l'inclusione, l'equità e la diversità culturale e linguistica nel suo impiego nei settori dell'istruzione e della ricerca. Tra le raccomandazioni principali figurano lo sviluppo di politiche nazionali per la protezione e la sicurezza dei dati, la definizione di criteri di validazione degli strumenti GenAI nelle istituzioni educative e accademiche, la promozione di iniziative formative per docenti, studenti e personale di staff delle università. Inoltre, nel documento si invita la comunità internazionale a riflettere sugli impatti di GenAI sia sull'insegnamento e sulla ricerca, sia sul modo in cui sono prodotte, diffuse e valutate la conoscenza e l'apprendimento.

Il Russell Group (2023) ha delineato una serie di raccomandazioni per guidare l'impiego di strumenti GenAI nelle università del Regno Unito appartenenti al loro gruppo. Le principali aree di intervento delle raccomandazioni riguardano l'uso appropriato e consapevole di strumenti GenAI, la formazione dei docenti e del personale di staff per un valido supporto degli studenti e l'adattamento dei metodi di insegnamento e di valutazione per favorire un utilizzo etico e l'implementazione di pratiche condivise tra le università del gruppo. In linea con tali principi, Russell Group mira a garantire che l'integrazione di GenAI avvenga in modo etico, responsabile e vantaggioso per gli stakeholder interni alle loro università.

Nel 2024 l'OCSE ha aggiornato la *Raccomandazione del Consiglio sull'Intelligenza artificiale*[5], adottata nel 2019, al fine di riflettere su alcuni sviluppi tecnologici (tra i quali GenAI) e politici. Il documento identifica cinque principi complementari importanti: crescita inclusiva, sviluppo sostenibile e benessere; rispetto dello stato di diritto, dei diritti umani e dei valori democratici; trasparenza e comprensibilità; robustezza, sicurezza e affidabilità; accountability. Sulla base di questi principi, si raccomanda agli Stati aderenti di: investire nella ricerca e nello sviluppo dell'IA; favorire lo sviluppo e l'accesso a un ecosistema digitale; creare un contesto di governance che favorisca l'IA; sviluppare le competenze umane e prepararsi al cambiamento del mercato del lavoro; collaborare globalmente per lo sviluppo di una IA

5 Si veda OCSE (2024), *Revised Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449, https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449.

affidabile. Particolare attenzione nel documento è riservata alla lotta alla disinformazione e alla tutela dell'integrità delle informazioni nell'era di GenAI.

Anche la Commissione europea ha recentemente pubblicato le *Living Guidelines on the responsible use of Generative AI in Research* (Commissione europea 2024) destinate a ricercatori, istituzioni accademiche e finanziatori pubblici e privati della ricerca. Le LG forniscono un quadro di riferimento comune per garantire l'uso appropriato ed efficace di GenAI, in risposta alla proliferazione di direttive provenienti dalle diverse istituzioni dell'ecosistema della ricerca. Il documento si basa su principi chiave della politica europea in materia di IA, come gli *Orientamenti etici per un'IA affidabile* (Commissione europea 2019), il *Codice di condotta europeo per l'integrità della ricerca* (ALLEA 2023) e l'AI Act, che attestano il ruolo guida che l'Unione europea vorrebbe assumere a livello mondiale su tale tecnologia. Le LG sottolineano sia il potenziale della tecnologia nel migliorare la produzione e la verifica delle conoscenze, sia i rischi di disinformazione e di cattiva condotta per quanto riguarda l'integrità scientifica. Il documento, inoltre, enfatizza i principi di affidabilità, onestà, rispetto e responsabilità, e sottolinea la necessità di un monitoraggio costante dell'evoluzione degli strumenti GenAI e delle implicazioni etiche e legali sul loro utilizzo.

Parallelamente alle iniziative di editori, organismi internazionali e istituzioni accademiche, molti ricercatori indagano l'impatto della regolamentazione dell'uso di strumenti GenAI nelle università attraverso l'analisi delle linee guida prodotte.

Bockting *et al.* (2023), su *Nature*, riconoscendo l'inevitabilità dello sviluppo e dell'uso di strumenti GenAI, spingono alla definizione di un quadro di governance per bilanciare il trade-off tra le opportunità offerte dalle innovazioni e i rischi connessi all'uso di tali strumenti. A tal fine, gli autori propongono la necessità di una supervisione scientifica che non può essere affidata esclusivamente a governi o aziende private, ma deve coinvolgere direttamente la comunità scientifica, attraverso l'istituzione di un organismo indipendente e interdisciplinare che monitori e certifichi gli strumenti di GenAI sulla base di criteri (accuratezza, bias e sicurezza ecc.) e garantisca sia l'accountability in ambito accademico e scientifico che la trasparenza sui dataset utilizzati. Infine, i ricercatori propongono un approccio dinamico e adattivo alla definizione delle LG, con meccanismi di aggiornamento regolare affiancati da strumenti di coinvolgimento attivo della comunità di riferimento (LG dinamiche), in modo che tale documento possa riflettere sia i progressi tecnici, sia l'emergere di nuove sensibilità etiche giuridiche e metodologiche legate all'impiego di strumenti GenAI.

Moorhouse *et al.* (2023), analizzando le LG delle prime 50 università nel *Times Higher Education World University Rankings* 2023, hanno riscontrato che meno della metà delle università hanno pubblicato LG relative all'uso di strumenti GenAI sui propri siti istituzionali. Gli autori hanno anche evidenziato la necessità di interventi a garanzia dell'integrità accademica, formando gli accademici nella comunicazione con gli studenti sul corretto utilizzo degli strumenti GenAI.

Smith *et al.* (2024), invece, basandosi sulle esperienze di due università australiane, propongono un quadro di riferimento pratico e adattabile che serva a: promuovere e facilitare l'uso responsabile di strumenti GenAI; orientare il complesso panorama normativo, alla luce dei rapidi cambiamenti di queste nuove tecnologie; definire delle LG basate su principi etici e di integrità.

McDonald *et al.* (2024) hanno esaminato documenti di policy e LG di 116 università statunitensi nel 2023, rilevando che il 63% di esse incoraggia l'uso di GenAI nell'insegnamento e nell'apprendimento. Gli autori hanno anche riscontrato che le LG analizzate pongono particolare enfasi sulla creazione di contenuti, sull'uso di strumenti GenAI, nonché su questioni di etica, di privacy e sull'impiego di strumenti per il rilevamento di lavori generati da GenAI.

Dabis e Csáki (2024), hanno individuato le principali dimensioni etiche rilevanti (responsabilità, supervisione umana e trasparenza) nelle politiche e nelle LG delle prime 30 università della *Shangai Global Ranking of Academic 2023*. Nel lavoro, gli autori raccomandano che le università sostengano una linea accademica rigorosa nell'affrontare le sfide poste dall'uso di strumenti GenAI promuovendo i suddetti principi etici e di trasparenza nei propri programmi e nei propri corsi.

Driessens e Pischetola (2024), utilizzando la metodologia di Bacchi sulla 'problematizzazione',

hanno esaminato le politiche sull'uso di strumenti GenAI adottate da 8 università danesi, evidenziando come l'attenzione sia prevalentemente focalizzata sull'integrità accademica, la privacy e la sicurezza dei dati, senza considerare pienamente le implicazioni politiche, economiche e di sostenibilità ambientale di tali strumenti per la didattica e la ricerca.

Yusuf *et al.* (2024), intervistando 1.217 studenti e docenti provenienti da 76 Paesi, dei quali oltre l'80% aveva un elevato livello di familiarità con gli strumenti GenAI, hanno evidenziato una forte correlazione tra le dimensioni culturali e le opinioni degli intervistati su benefici e rischi relativi all'uso di GenAI. Gli autori sostengono che un uso responsabile di strumenti GenAI, attraverso la definizione di LG sviluppate su principi etici, può migliorare i processi di apprendimento e di ricerca; ma per quanto riguarda gli aspetti legati alla multiculturalità sono necessarie delle politiche di intervento appropriate che bilancino innovazione e integrità accademica, considerando la presenza di molteplici contesti culturali che si approcciano differentemente all'uso di strumenti GenAI.

Ullah *et al.* (2024) hanno esaminato le LG pubblicate dalle prime 50 università del *QS World University Ranking* 2025. Nelle 41 LG disponibili emerge l'attenzione verso un uso appropriato e responsabile degli strumenti GenAI. Un elemento ricorrente riguarda le limitazioni all'uso di tale tecnologia spesso legate alla preventiva autorizzazione da parte del docente/supervisore. Le LG analizzate, infine, forniscono anche indicazioni specifiche sull'uso etico e responsabile degli strumenti GenAI in aula e nella valutazione degli studenti.

Per concludere, la letteratura evidenzia che l'uso di strumenti GenAI rappresenta una straordinaria opportunità per l'accademia e la ricerca scientifica, in quanto consente di accelerare la produzione e la diffusione della conoscenza. Tuttavia, tutto ciò necessità di un quadro normativo solido, di una supervisione umana rigorosa e di LG chiare ed efficaci che garantiscano qualità, affidabilità e integrità nei processi di ricerca. Un approccio responsabile e trasparente che orienti lo sviluppo di strumenti GenAI verso un equilibrio tra l'esplorazione del suo potenziale e il rispetto degli standard di validità, originalità, riproducibilità e veridicità scientifica.

2. Metodologia

La prima fase del lavoro è stata dedicata all'identificazione delle università europee presenti nei primi 50 posti dei tre prestigiosi ranking delle migliori università nel mondo: *Quacquarelli Symonds (QS) World University Ranking 2025*, *Times Higher Education World University Rankings* 2025, *Shangai Global Ranking of Academic 2024*[6].

A seguito di un lavoro di pulizia dei dati, dovuto alla presenza di molte università su più ranking, sono state individuate 20 università, riportate nella tabella 1.

Nell'ambito dell'analisi svolta tra febbraio e marzo 2025, gli Autori hanno verificato la presenza di LG disponibili sui siti web di 10 delle università presenti nella tabella 1. La seconda fase del lavoro è stata dedicata alla definizione della checklist per l'analisi delle suddette LG, costruita in base alla letteratura riportata nel paragrafo 1. La suddetta checklist è stata testata in via sperimentale sulle prime 3 università presenti nella tabella 1. Tale attività ha apportato delle piccole modifiche, a seguito delle quali è stata ricavata la checklist definitiva composta da 18 criteri suddivisi in 4 categorie: *Presenza delle LG sui siti web delle università; Informazioni generali presenti nelle LG sull'uso di strumenti GenAI; Istruzioni presenti nelle LG sull'uso di strumenti GenAI; Istruzioni presenti nelle LG su principi etici e aspetti legali sull'uso di strumenti GenAI* (si veda Appendice 2).

In questo studio è stato utilizzato un metodo quantitativo per esaminare i contenuti delle suddette LG, alla luce dei 18 criteri della checklist. Anche questa fase di lavoro si è conclusa con un'operazione di pulizia del dato per renderlo maggiormente fruibile per l'analisi dei risultati e delle evidenze.

6 Si vedano in particolare: Quacquarelli Symonds (2025), *QS World University Ranking 2025*, https://www.topuniversities.com/world-university-rankings; Times Higher Education (2025), *The Times Higher Education World University Rankings 2025,* https://www.timeshighereducation.com/world-university-rankings/latest/world-ranking; Shanghai Academic Ranking of World Universities (2025), *Global Ranking of Academic 2024*, https://www.shanghairanking.com/rankings/arwu/2024 (ultimo accesso ai siti web 19 marzo 2025).

Tabella 1. Elenco delle università europee presenti nei tre ranking

Università	Presenza nei ranking*
Imperial College London	1,2,3
University of Oxford	1,2,3
University of Cambridge	1,2,3
ETH Zurich - Swiss Federal Institute of Technology	1,2,3
University College London (UCL)	1,2,3
Université PSL (Paris Sciences & Lettres)	1,2,3
The University of Edinburgh	1,2,3
Technical University of Munich (TUM)	1,2,3
École Polytechnique Fédérale de Lausanne (EPFL)	1,2
King's College London	1,2
Ludwig-Maximilians-Universität München (LMU)	2,3
Heidelberg University	2,3
Karolinska Institute	2,3
Institut Polytechnique de Paris	1
Delft University of Technology	1
KU Leuven	2
London School of Economics and Political Science	2
Paris-Saclay University	3
University of Copenhagen	3
Sorbonne University	3

Nota: *Legenda della presenza delle università nei ranking 1=Quacquarelli Symonds (QS) World University Ranking 2025; 2=Times Higher Education World University Rankings 2025; 3=Shangai Global Ranking of Academic 2024.

Fonte: elaborazione degli Autori, 2025

3. Risultati

Categoria: Presenza delle LG sui siti web delle università

Per quanto riguarda tale categoria, l'analisi svolta ha evidenziato la disponibilità al pubblico di LG sui siti web di 10 università (50%), riportate in Appendice 1 con i relativi URL per visionarle (ultima consultazione 19 marzo 2025). Delle altre 10 università in cui non sono disponibili LG, si evidenzia:

- *Karolinska Institute* dichiara sul proprio sito web che non hanno delle LG ma consigli e suggerimenti (*Generative AI and teaching - Advice for educators*) redatti dalla loro *Unit for Teaching and Learning* (UoL);
- *Ludwig-Maximilians-Universität München (LMU)* ha sul proprio sito web il documento *Guidelines on writing seminar papers, bachelor and master theses* (LMU Munich School of Management Professorship for International Management, Prof. Dr. Helene Tenzer), il cui paragrafo 6, *Using AI in Academic Writing,* riporta indicazioni sull'uso di strumenti GenAI per la redazione di paper per seminari e tesi di laurea triennale e magistrale.

Categoria: Informazioni generali presenti nelle LG sull'uso di strumenti GenAI

Per quanto riguarda tale categoria, l'analisi svolta ha prodotto i risultati riportati in ordine di graduatoria nella tabella 2.

Come si può notare, tutte le università (100%) dichiarano i loro obiettivi e gli ambiti di intervento nelle loro LG. La maggior parte di esse (90%) ha indicato l'autorità che ha emesso tali LG: gli uffici degli Organi di Ateneo (30%); i centri/servizi informativi e tecnologici (30%); i centri di insegnamento e apprendimento (20%); la biblioteca universitaria (10%). Il 90% delle università ha indicato gli esempi di strumenti GenAI presenti nelle loro LG. Microsoft Copilot è stato menzionato da 8 università (80%), mentre ChatGPT solo da 4 (40%). Le università hanno menzionato in totale 12 strumenti GenAI specifici

Tabella 2. Presenza (sì/no) di informazioni generali presenti sulle LG (n=10)

Ordine di graduatoria	Criteri	SÌ	NO
1	Obiettivi e scopi delle LG	100%	0%
2-3	Autorità che ha emesso le LG	90%	10%
2-3	Esempi di strumenti GenAI presenti nelle LG delle università	90%	10%
4	Data di rilascio/ultimo aggiornamento	80%	20%
5	Informazioni di contatto per le LG	60%	40%

Fonte: elaborazione degli Autori

come elaboratori di testi e di immagini. L'80% delle università ha indicato la data di rilascio o di ultimo aggiornamento delle LG che, alla data ultima di consultazione dei siti web (19 marzo 2025), per tutti gli atenei è l'anno 2024. Solo il 60% delle università ha indicato indirizzi ed e-mail dei contatti per le LG; due università hanno indicato anche i nominativi dei referenti (20%). In un solo caso è stata riportata la persona che ha realizzato le LG, non indicandola come referente.

Categoria: Istruzioni presenti nelle LG sull'uso di strumenti GenAI

Per quanto riguarda tale categoria, l'analisi svolta ha prodotto i risultati riportati in ordine di graduatoria nella tabella 3.

Tutte le 10 università hanno dichiarato nelle LG che è consentito l'uso di strumenti GenAI (100%); stessa percentuale è stata riscontrata anche per quanto riguarda l'uso non consentito di strumenti GenAI (100%). Il 70% delle università prevede l'opportunità, in alcuni casi anche l'obbligo, di dover chiedere l'approvazione del docente/supervisore all'utilizzo di strumenti GenAI. Il 60% delle università riporta le istruzioni per la dichiarazione di uso di strumenti GenAI (descrizione/nome/versione/data). A tale dichiarazione, il 50% delle università associa anche degli esempi pratici su come tale descrizione deve essere riportata nei documenti in cui tali strumenti sono stati utilizzati. Alcune università (30%) richiedono che nei suddetti documenti siano riportati anche i prompt utilizzati per i contenuti generati. Il 30% delle università esige che l'output generato da strumenti GenAI sia modificato e migliorato dagli utenti in base alle indicazioni fornite nelle LG. Per quanto riguarda la corretta integrazione e l'uso di strumenti GenAI in aula e nella valutazione, il 30% delle università indica le modalità di utilizzo. Invece, per favorire l'uso corretto di strumenti GenAI, il 60% delle università prevede corsi di formazione iniziale e di aggiornamento per migliorare le conoscenze e le competenze, erogati in presenza o da remoto, e in varie modalità (audio, video ecc.) per apportare valore aggiunto alla didattica e alla ricerca.

Tabella 3. Presenza (sì/no) di istruzioni nelle LG (n=10)

Ordine di graduatoria	Criteri	SÌ	NO
1-2	Utilizzo consentito di strumenti GenAI	100%	0%
1-2	Utilizzo non consentito di strumenti GenAI	100%	0%
3	Approvazione del docente/supervisore	70%	30%
4-5	Istruzioni per la dichiarazione di uso di strumenti GenAI (descrizione/nome/versione/data)	60%	40%
4-5	Formazione e aggiornamento continuo di conoscenze e competenze di strumenti GenAI	60%	40%
6	Presenza di esempi nella dichiarazione di uso di strumenti GenAI	50%	50%
7-9	Informazioni (dettagli) sui prompt di strumenti GenAI	30%	70%
7-9	Istruzioni sull'uso e l'adattamento degli output di strumenti GenAI	30%	70%
7-9	Istruzioni per l'utilizzo di strumenti GenAI in aula e valutazioni	30%	70%

Fonte: elaborazione degli Autori

Tabella 4. Presenza (sì/no) di istruzioni su principi etici e aspetti legali nelle LG (n=10)

Ordine di graduatoria	Criteri	SÌ	NO
1-2	Utilizzo di strumenti GenAI in modo trasparente	90%	10%
1-2	Attenzione alla privacy, riservatezza, sicurezza dei dati e diritti di proprietà intellettuale	90%	10%
3	Attenzione all'integrità accademica e alla cattiva condotta (Responsabilità finale dell'output GenAI, riferimento e citazione degli output GenAI, presenza di iter di cattiva condotta)	50%	50%

Fonte: elaborazione degli Autori

Categoria: Istruzioni presenti nelle LG su principi etici e aspetti legali sull'uso di strumenti GenAI

Per quanto riguarda tale categoria, l'analisi svolta ha prodotto i risultati riportati nella tabella 4.

Il rispetto della trasparenza è rivolto soprattutto alla valutazione degli output, evidenziata dal 90% delle università. Su questo tema c'è particolare attenzione nelle LG dovuta al timore che gli stakeholder interni dell'università, tra i quali soprattutto gli studenti, non prestino la massima attenzione ai risultati degli output e al loro utilizzo. Attenzione che resta alta per quanto riguarda gli aspetti di privacy, sicurezza dei dati e copyright (90%).

Solo il 50% delle università evidenzia l'attenzione all'integrità accademica, ciò è dovuto in quanto tale criterio è associato alla cattiva condotta; infatti, non tutte le università riportano nelle loro LG un iter procedurale nel caso di cattivi comportamenti, soprattutto da parte degli studenti.

4. Principali evidenze

Dall'ingresso di ChatGPT molte università sono state impegnate nella definizione di LG: da un lato, per garantire un uso trasparente di strumenti di GenAI in linea con i principi etici e di integrità della ricerca; dall'altro, per supportare docenti, studenti e personale di staff nella costruzione di un'alfabetizzazione critica e un uso appropriato di questa nuova tecnologia. Numerosi lavori stanno indagando un ambito ancora poco esplorato in letteratura quale la presenza di LG che regolamentino l'uso di strumenti GenAI nelle università europee. LG che richiedono un impegno continuo di revisione e aggiornamento per rimanere al passo con gli strumenti GenAI, che si innovano in modo rapido e sorprendente (Cacho 2024).

Il presente lavoro ha rilevato che nelle prime 20 università europee, presenti nelle prime 50 posizioni dei tre ranking delle migliori università del mondo, solo 10 (50%) rendono disponibili sui loro siti web specifiche LG sull'uso di strumenti GenAI.

Sempre a livello generale si evidenzia che le LG di tutte le università britanniche presenti in tabella 1 si rifanno a *Russell Group principles on the use of generative AI tools in education*; mentre EPFL e KU Luven si rifanno alle *Living guidelines on the responsible use of generative AI in research* della Commissione europea.

Il King's college, diversamente dalle altre università, rende disponibili al pubblico sul proprio sito web una serie di LG (*King's guidance on generative AI for teaching, assessment and feedback*[7]): *Macro-level guidance* (University-wide principles and policy); *Meso-level guidance* (Departments, programmes and modules); *Micro-level guidance* (GTAs/teachers/lectures/PhD supervisors); *Approaches to assessment in the age of AI* (Support for staff in making changes to their assessment approaches); *Student guidance* (Student guidance on the use of generative AI); *Guidance for doctoral students, supervisors and examiners* (Permitted uses of generative AI tools in doctoral assessment); *PAIR (problem, AI, interaction, reflection) framework guidance* (Proactively integrating AI into our curriculum).

Nel dettaglio, l'analisi dei contenuti delle linee guida ha evidenziato che tutte riportano in modo esplicito gli obiettivi e gli ambiti di intervento di riferimento; nonché una presentazione generale di GenAI (con relativa spiegazione del suo funzionamento), corredata dall'esposizione di benefici e soprattutto rischi derivanti dall'uso di strumenti GenAI (pregiudizi, allucinazioni, *black box* ecc.) al fine di un loro corretto utilizzo nelle attività universitarie e nella ricerca in generale.

Nella maggioranza delle LG sono indicate l'autorità che le ha emesse, la data di rilascio o

7 Si veda https://www.kcl.ac.uk/about/strategy/learning-and-teaching/ai-guidance.

di ultimo aggiornamento e, in buona parte di esse, riferimenti ed e-mail di contatto. Il 90% delle università ha indicato i nomi di strumenti GenAI presenti nelle loro LG. Su questo punto si evidenzia che, diversamente da altri studi presenti in letteratura (Ghimire e Edwards 2024; Yusuf *et al.* 2024; Ullah *et al.* 2024), è Microsoft Copilot e non ChatGPT lo strumento GenAI più menzionato, in linea con quanto affermato in letteratura che la popolarità e l'uso di tali strumenti sono legati alle diverse discipline, università e Paesi (Lee *et al.* 2024).

Tutte le università non solo consentono, ma spingono gli stakeholder interni, in particolare gli studenti, all'utilizzo di strumenti GenAI, nelle more di un uso etico e appropriato delle stesse in un'ottica di salvaguardia dell'integrità della ricerca. In questa direzione, le LG riportano esempi in cui è consentito l'utilizzo di tali strumenti: fornire una panoramica di nuovi concetti e agire come coach collaborativo per risolvere i problemi; reinterpretare le informazioni per presentare modi alternativi di esprimere idee (appunti personali delle lezioni ecc.); supportare la gestione del tempo e suggerire metodi per affrontare compiti complessi; generare idee o superare il blocco dello scrittore; dare un senso a informazioni complesse; sollecitare domande riflessive per facilitare l'apprendimento ecc. Di conseguenza, tutte le università riportano le limitazioni, se non i divieti, all'uso di strumenti GenAI al fine di proteggere i dati personali e la sicurezza degli studenti ed evitare condotte accademiche scorrette, quali: sostenere test, scrivere relazioni di ricerca, imparare cose nuove o cercare informazioni, generare contenuti di cui non si è in grado di controllarne la veridicità o la forma, creare un codice informatico, completare i principali incarichi del corso, nonché in tutte le attività in cui tale uso non è esplicitato.

Il 70% delle università prevede la possibilità, in alcuni casi anche l'obbligo, di dover chiedere l'approvazione del docente/supervisore all'utilizzo di strumenti GenAI, soprattutto nei casi in cui non si è sicuri di come utilizzarli in modo corretto. Buona parte delle università (60%) definisce le istruzioni per la dichiarazione di uso di strumenti GenAI (descrizione/nome/versione/data); il 50% di esse aggiunge anche degli esempi pratici a tale dichiarazione, come nel caso riportato di seguito tratto dalle LG di Imperial college: *La dichiarazione dovrebbe essere scritta in frasi complete e includere le seguenti informazioni: nome e versione dello strumento di intelligenza artificiale generativa utilizzato, come ad esempio ChatGPT-3.5; editore (nome dell'azienda che fornisce il sistema di intelligenza artificiale), ad esempio OpenAI URL dello strumento di intelligenza artificiale; breve descrizione (frase singola) del modo in cui è stato utilizzato lo strumento; conferma che il lavoro sia tuo, ad esempio: riconosco l'utilizzo di ChatGPT-3.5 (OpenAI, https://chat.openai.com/) per generare uno schema per lo studio di base*. Circa un terzo delle università richiedono anche informazioni sui prompt utilizzati per i contenuti generati, non solo in un'ottica di trasparenza, ma anche per un'utilità derivante dalla riproducibilità del lavoro prodotto. Tali informazioni devono essere riportate alla fine dei documenti realizzati, anche se ciò potrebbe procurare un aggravio di lavoro per lo scrivente. Il 30% delle università indica nelle LG le istruzioni sull'uso e l'adattamento degli output prodotti da strumenti GenAI, in modo da evitare che gli stessi siano 'incollati' direttamente nei lavori senza che siano apportate modifiche, correzioni e migliorie da studenti e altri stakeholder interni. Il 30% delle università evidenzia nelle LG anche l'uso e la corretta integrazione di strumenti GenAI in aula e nella valutazione, in quanto riconosce il valore aggiunto dell'utilizzo di tali strumenti, come ad esempio nella pianificazione dei programmi di studio, nella creazione di contenuti, nelle pratiche convenzionali di insegnamento e nella fase di assesment. Per incentivare un uso corretto di strumenti GenAI, il 60% delle università propone programmi di formazione e di aggiornamento delle conoscenze di tali strumenti per utilizzarli al meglio nella didattica e nella ricerca.

Infine, in tutte le LG analizzate, gli autori, al di là di sporadiche citazioni del Regolamento (UE) 2016/679, non hanno riscontrato la presenza di riferimenti legislativi nazionali, europee e internazionali.

Sui temi della privacy, della sicurezza dei dati, della proprietà intellettuale e dell'integrità accademica, tutte le università sostengono una linea accademica rigorosa di protezione di tali principi, confermando quanto riportato in letteratura (Dabis e Csàki 2024). Il rispetto della trasparenza è rivolto soprattutto alla valutazione degli output, richiamata nella quasi totalità delle LG esaminate (90%), nelle quali si chiede una valutazione critica, ma soprattutto una verifica delle informazioni generate dagli

strumenti GenAI. Molte LG richiamano un'attenzione alle forme di plagio, anche se sono poche quelle che fanno riferimenti a *tools* di rilevamento di tale pratica scorretta. Ciò dovuto, ed è anche conclamato nelle LG, alla scarsa efficacia e affidabilità di questi strumenti nell'identificare in modo indiscutibile contenuti generati da GenAI. Infatti, in quasi tutte le LG, più che percorrere tale strada, si preferisce far appello ai principi etici di integrità e rigore accademico.

Conclusioni

Le LG esaminate in questo studio confermano che nelle università gli strumenti GenAI possono fornire un valido supporto per la didattica, la produzione accademica e la gestione amministrativa; anche se il loro utilizzo deve essere regolamentato da principi etici, di trasparenza e di rispetto della sicurezza e della privacy dei dati.

Un aspetto chiave, emerso anche in questo lavoro, è la necessità di promuovere l'alfabetizzazione e l'aggiornamento delle conoscenze e competenze di studenti, docenti, ricercatori e personale di staff all'uso di strumenti GenAI. Le istituzioni accademiche, infatti, non devono solo limitarsi a stabilire politiche di condotta o definire LG, ma devono promuovere programmi di formazione che permettano di comprendere appieno il funzionamento, le limitazioni e le implicazioni dovute all'uso di questi strumenti. Tutto ciò al fine di evitare che si perpetuino disuguaglianze esistenti e favorire l'inclusione, soprattutto di gruppi svantaggiati come, ad esempio, studenti internazionali (non di madre lingua), non nativi digitali e persone con disabilità.

Per il futuro, sarà fondamentale sviluppare LG dinamiche e aggiornabili in risposta ai rapidi progressi tecnologici e alle nuove sfide emergenti indotte dall'uso di strumenti GenAI. Le istituzioni accademiche dovrebbero collaborare a livello internazionale per armonizzare le politiche di condotta e definire standard condivisi che garantiscano un utilizzo responsabile ed equo di strumenti GenAI. In questa direzione si evidenzia la proposta avanzata da qualche autore sulla costituzione di un organismo internazionale autonomo che definisca gli standard e i framework a cui le LG dovrebbero fare riferimento (Bockting *et al.* 2023). I futuri lavori, inoltre, dovranno concentrarsi sugli effetti di lungo periodo di queste tecnologie sulla didattica e sulla ricerca, individuando strategie per mitigare i rischi e massimizzare i benefici.

Infine, l'integrazione di GenAI nella produzione scientifica solleva questioni etiche legate all'attribuzione degli autori e alla qualità dei contenuti generati. Affinché GenAI diventi una tecnologia affidabile e accettata nella comunità accademica, è essenziale sviluppare politiche chiare sull'uso di questi strumenti nella scrittura e nella revisione scientifica. Solo un approccio incentrato sull'uomo, ispirato ai principi etici delineati dall'Unesco, dall'UE e da altre organizzazioni internazionali, può garantire che GenAI contribuisca a un ecosistema della ricerca più efficiente, inclusivo e innovativo.

Per concludere, il presente lavoro presenta alcune limitazioni nell'analisi, nell'interpretazione e nella presentazione dei dati, che potranno essere affrontate in futuri lavori. Di seguito si evidenziano le più rilevanti.

Pur appartenendo allo stesso continente e figurando tra le prime 50 posizioni nei tre ranking considerati, le università del presente lavoro differiscono per cultura organizzativa, dotazione tecnologica e priorità strategiche, tali variabili dovrebbero essere analizzate in modo più approfondito.

Altri aspetti da esaminare sono l'influenza dei contesti istituzionali e della multiculturalità, derivante dalla composizione internazionale di docenti e studenti, sull'utilizzo delle LG. A ciò si aggiunge la diversa normativa a livello nazionale, ovvero i quadri regolatori e giuridici (ad es. trasparenza, privacy, sicurezza dei dati ecc.), e i principi etici che variano da paese a paese, influenzando l'impianto delle politiche e delle stesse linee guida. In ultimo, l'analisi potrebbe essere stata parziale in quanto, anche se le linee guida di 10 università non erano disponibili al pubblico sui siti web, ciò non vuol dire che tali università non le abbiano emanate.

Appendice 1

Lista delle università europee presenti nei tre ranking e URL delle loro line guida (consultazione febbraio-marzo 2025)

Università	URL LG
Imperial College London	https://www.imperial.ac.uk/admin-services/library/learning-support/generative-ai-guidance/
University of Oxford	https://communications.admin.ox.ac.uk/communications-resources/ai-guidance
University of Cambridge	https://blendedlearning.cam.ac.uk/guidance-support/ai-and-education/using-generative-ai
ETH Zurich - Swiss Federal Institute of Technology	https://ethz.ch/staffnet/en/news-and-events/internal-news/archive/2024/07/new-guidelines-for-the-use-of-generative-ai-in-education.html
University College London (UCL)	https://www.ucl.ac.uk/teaching-learning/publications/2023/sep/using-generative-ai-genai-learning-and-teaching
The University of Edinburgh	https://information-services.ed.ac.uk/computing/comms-and-collab/elm/guidance-for-working-with-generative-ai https://information-services.ed.ac.uk/computing/comms-and-collab/elm/generative-ai-guidance-for-staff
École Polytechnique Fédérale de Lausanne (EPFL)	https://www.epfl.ch/about/vice-presidencies/vice-presidency-for-academic-affairs-vpa/tips-for-the-use-of-generative-ai-in-research-and-education/
King's College London	https://www.kcl.ac.uk/about/strategy/learning-and-teaching/ai-guidance
KU Leuven	https://www.kuleuven.be/english/genai/general-guidelines
London School of Economics and Political Science	https://info.lse.ac.uk/staff/divisions/Eden-Centre/Artificial-Intelligence-Education-and-Assessment/School-position-on-generative-AI
Karolinska Institute	Non sono presenti LG sul sito web Sul sito web, l'Università dichiara che non ha LG ma solo consigli e suggerimenti (*Generative AI and teaching - Advice for educators*) forniti da Unit for Teaching and Learning (UoL). https://staff.ki.se/education-support/teaching-and-learning/generative-ai-and-teaching-advice-for-educators
Ludwig-Maximilians-Universität München (LMU)	Non sono presenti LG sul sito web Sul sito web è disponibile il documento *"Guidelines on writing seminar papers, bachelor and master theses"*, LMU Munich School of Management Professorship for International Management, Prof. Dr. Helene Tenzer, il cui paragrafo "*6. Using AI in Academic Writing*" riporta indicazioni sull'uso di strumenti GenAI per la redazione di papers per seminari e tesi di laurea triennale e magistrale.
Technical University of Munich (TUM)	Non sono presenti LG sul sito web
Heidelberg University	Non sono presenti LG sul sito web
Institut Polytechnique de Paris	Non sono presenti LG sul sito web
Delft University of Technology	Non sono presenti LG sul sito web
Université PSL (Paris Sciences & Lettres)	Non sono presenti LG sul sito web
Paris-Saclay University	Non sono presenti LG sul sito web
University of Copenhagen	Non sono presenti LG sul sito web
Sorbonne University	Non sono presenti LG sul sito web

Fonte: elaborazione degli Autori

Appendice 2

Strumento per la valutazione delle LG sull'uso di strumenti di intelligenza artificiale generativa nelle università europee

Categorie	Check list	Sì	NO
Presenza delle LG sui siti web delle università	Presenza o assenza delle LG disponibili al pubblico sui siti web		
Informazioni generali presenti nelle LG sull'uso di strumenti GenAI	1. Obiettivi e scopi delle LG		
	2. Autorità che ha emesso le LG		
	3. Data di rilascio/ultimo aggiornamento		
	4. Informazioni di contatto per le LG		
	5. Esempi di strumenti GenAI presenti nelle LG delle università		
Istruzioni presenti nelle LG sull'uso di strumenti GenAI	6. Utilizzo consentito di strumenti GenAI		
	7. Utilizzo non consentito di strumenti GenAI		
	8. Approvazione del docente/supervisore		
	9. Istruzioni per la dichiarazione di uso di strumenti GenAI (descrizione/ nome/versione/data)		
	10. Presenza di esempi della dichiarazione di uso di strumenti GenAI		
	11. Informazioni (dettagli) sui prompt di strumenti GenAI		
	12. Istruzioni sull'uso e l'adattamento degli output di strumenti GenAI		
	13. Istruzioni per l'utilizzo di strumenti GenAI in aula e valutazioni		
	14. Formazione e aggiornamento continuo di conoscenze e competenze di strumenti GenAI		
Istruzioni presenti nelle LG sui principi etici e aspetti legali sull'uso di strumenti di strumenti GenAI	15. Attenzione alla privacy, riservatezza, sicurezza dei dati e diritti di proprietà intellettuale		
	16. Utilizzo di strumenti GenAI in modo trasparente		
	17. Attenzione all'integrità accademica e alla cattiva condotta (Responsabilità finale dell'output GenAI, riferimento e citazione degli output GenAI, presenza di iter di cattiva condotta)		

Fonte: elaborazione degli Autori

Bibliografia

Agomuoh F. (2023), Professor flunks all his students after ChatGPT false claims*, digitaltrends.com*, May 18

Aljamaan F., Temsah M.H., Altamimi I., Al-Eyadhy A., Jamal A., Alhasan K., Mesallam T.A., Farahat M., Malki K.H. (2024), Reference Hallucination Score for Medical Artificial Intelligence Chatbots: Development and Usability Study, *JMIR Medical Informatics*, 12 <DOI: 10.2196/54345>

ALLEA (2023), *The European Code of Conduct for Research Integrity. Revised Edition*, Berlin, ALLEA- All European Academies

Amir O., Kamar E., Kolobov A., Grosz B.J. (2016), Interactive Teaching Strategies for Agent Training, in Kambhampati S. (ed.), *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)*, Palo Alto California USA, AAAI Press International Joint Conferences on Artificial Intelligence, pp.804-811

Antonenko P., Abramowitz B. (2023), In-service teachers'(mis)conceptions of artificial intelligence in K–12 science education, *Journal of Research on Technology in Education*, 55, n.1, pp.64-78

Aspen Institute Italia (2024), *Rapporto Intelligenza artificiale 2024*, Roma, Aspen Institute Italia, Osservatorio Permanente sull'Adozione e l'Integrazione della Intelligenza artificiale

Baidoo-Anu D., Ansah L.O. (2023), Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning, *Journal of AI*, 7, n.1, pp.52-62

Bathaee Y. (2018), The artificial intelligence black box and the failure of intent and causation, *Harvard Journal of Law & Technology*, 31, n.2, pp.889-938

Bockting C.L., van Dis E.A.M., van Rooij R., Zuidema W., Bollen J. (2023), Living Guidelines for Generative AI in Research, *Nature*, 622, pp.693-696

Brown T.B., Mann B., Ryder N., Subbiah M., Kaplan J., Dhariwal P., Neelakantan A., Shyam P., Sastry G., Askell A., Agarwal S., Herbert-Voss A., Krueger G., Henighan T., Child R., Ramesh A., Ziegler D.M., Wu J., Winter C., Hesse C., Chen M., Sigler E., Litwin M., Gray S., Chess B., Clark J., Berner C., McCandlish S., Radford A., Sutskever I., Amodei D (2020), *Language models are few-shot learners*, 34th International Conference on Neural Information Processing Systems, NeurIPS, Curran Associates Inc.

Cacho R. (2024), Integrating Generative AI in University Teaching and Learning: A Model for Balanced Guidelines, ***Online Learning***, 28, n.3, pp.55-81

Chan C.K.Y. (2023), A comprehensive AI policy education framework for university teaching and learning, *International Journal of Educational Technology in Higher Education,* 20, n.38 <https://doi.org/10.1186/s41239-023-00408-3>

Commissione europea (2024), *Living guidelines on the responsible use of generative AI in research. ERA Forum Stakeholders'document*, Brussels, European Union

Commissione europea (2019), *Orientamenti etici per un'IA affidabile*, Bruxelles, Commissione europea <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

COPE (2023), *Authorship and AI tools: COPE position statement*, Committee on Publication Ethics <Doi.org/10.24318/cCVRZBms>

Cornell University Task Force (2023), *Generative AI in Academic Research: Perspectives and Cultural Norms*, Ithaca (NW), Cornell University

CRFM (2022), *On the opportunities and risks of foundation models,* Stanford, Center for Research on Foundation Models-Stanford Institute for Human-Centered Artificial Intelligence-Stanford University

Dabis A., Csáki C. (2024), AI and ethics: Investigating the first policy responses of higher education institutions to the challenge of generative AI, *Humanities and Social Sciences Communications*, 11, n.1006 <Doi.org/10.1057/s41599-024-03526-z>

Driessens O., Pischetola M. (2024), Danish university policies on generative AI: Problems, assumptions and sustainability blind spots, *MedieKultur: International Journal of Media and Communication Research*, 40, n.76, pp.31-52

Eaton S.E. (2023), Postplagiarism: transdisciplinary ethics and integrity in the age of artificial intelligence and neurotechnology, *International Journal for Educational Integrity,* 19, n.23 <https://Doi.org/10.1007/s40979-023-00144-1>

Floridi L. (2025), *Distant Writing: Literary Production in the Age of Artificial Intelligence*, Centre for Digital Ethics (CEDE) Research Paper <http://dx.Doi.org/10.2139/ssrn.5232088>

Ghimire A., Edwards J. (2024), *From guidelines to governance: A study of ai policies in education*, in 25th International Conference, AIED 2024, Recife, Brazil, July 8-12, Proceedings, Part II, <https://Doi.org/10.48550/arXiv.2403.15601>

Halaweh M. (2023), ChatGPT in education: Strategies for responsible implementation. *Contemporary Educational Technology*, 15, n.2, ep421

Hutson J. (2024), The Evolving Role of Copyright Law in the Age of AI-Generated Works, *Journal of Digital Technologies and Law,* 2, n.4, pp.886-914

Jiang Z., Guo M., Hu Y., Gong N.Z. (2024), *Watermark-based Detection and Attribution of AI-Generated Content*, Durham (NC), Duke University

Kaplan-Rakowski R., Grotewold K., Hartwick P., Papin K. (2023), Generative AI and Teachers' Perspectives on Its Implementation in Education, *Journal of Interactive Learning Research,* 34, n.2, pp.313-338

Khalil M., Er E. (2023), Will ChatGPT Get You Caught? Rethinking of Plagiarism Detection, *arxiv.org* <https://Doi.org/10.48550/arXiv.2302.04335>

Lee D., Arnold M., Srivastava A., Plastow K., Strelan P., Ploeckl F., Lekk D., Palmere E. (2024), The impact of generative AI on higher education learning and teaching: A study of educators'perspectives, *Computer and Education: Artificial Intelligence*, 6, n.100221

Miao F., Holmes W. (2023), *Guidance for Generative AI in Education and Research*, Paris, UNESCO Publishing

McDonald N., Johri A., Ali A., Hingle A. (2024), Generative artificial intelligence in higher education: Evidence from an analysis of institutional policies and guidelines, *arxiv.org* <https://Doi.org/10.48550/arXiv.2402.01659>

Moorhouse B.L., Yeo M.A., Wan Y. (2023), Generative AI tools and assessment: Guidelines of the world's top-ranking universities, *Computers and Education Open*, 5, n.2, 100151

Nature (2023), Tools such as ChatGPT threaten transparent science; here are our ground rules for their use, *Nature editorial*, 613, n.7945

Osoba O.A., Welser W. (2017), *An Intelligence in Our Image. The Risks of Bias and Errors in Artificial Intelligence*, Santa Monica (CA) RAND Corporation

Russell Group (2023), *Russell Group Principles on the Use of Generative AI Tools in Education*, Cambridge (UK), Russell Group

Sætra H.S. (2020), A shallow defence of a technocracy of artificial intelligence: Examining the political harms of algorithmic governance in the domain of government, *Technology in Society*, 62, n.101283

Smith S., Tate M., Freeman K., Walsh A., Ballsun-Stanton B., Hooper M., Lane M. (2024), A university framework for the responsible use of generative AI in research, *arxiv.org* <https://doi.org/10.48550/arXiv.2404.19244>

Stokel-Walker C. (2023), ChatGPT listed as author on research papers: many scientists disapprove, *Nature*, 613, pp.620-621

Ullah M., Bin Naeem S., Kamel Boulos M.N. (2024), Assessing the Guidelines on the Use of Generative Artificial Intelligence Tools in Universities: A Survey of the World's Top 50 Universities, *Big Data and Cognitive Computing*, 8, n.12, 194

von Krogh G., Roberson Q., Gruber M. (2023), Recognizing and utilizing novel research opportunities with artificial intelligence, *Academy of Management Journal,* 66, n.2, pp.367-373

Yusuf A., Pervin N., Román-González M. (2024), Generative AI and the future of higher education: A threat to academic integrity or reformation? Evidence from multicultural perspectives, *International Journal of Educational Technology in Higher Education*, 21, n.12 <DOI:10.1186/s41239-024-00453-6>

Filippo Giordano
f.giordano@lumsa.it
Professore ordinario di Economia aziendale e direttore del Dipartimento di Giurisprudenza, economia, politica e lingue moderne dell'Università LUMSA di Roma. Dal 2022 è Chair del SIG di Public and Nonprofit Management di EURAM – European Academy of Management. È docente di Business Ethics presso l'Università Bocconi e research affiliate presso il centro ICRIOS Bocconi. Dal 2004 è docente della SDA Bocconi School of Management.

Saverio Lovergine
s.lovergine@inapp.gov.it
PhD, è Funzionario in Inapp ed è componente dell'Unità centrale del Dipartimento della Funzione pubblica della Presidenza del Consiglio per la riforma del mercato del lavoro della PA, nonché senior consultant di numerose AA.PP. Rappresenta lo Stato italiano in ESCO mswg in Commissione europea. Collabora con la cattedra di Economia aziendale dell'Università di Roma "Tor Vergata" ed è componente del comitato editoriale di *European Scientific Journal*.

La progettazione sociale e la sfida dell'Intelligenza artificiale

Alberto Bortolami
Associazione italiana progettisti sociali APIS

Francesco Capone
Associazione italiana progettisti sociali APIS

Giusy Caravello
Associazione italiana progettisti sociali APIS

Antonio Finazzi Agrò
Associazione italiana progettisti sociali APIS

Federico Maggiora
Associazione italiana progettisti sociali APIS

Annaleda Mazzucato
Associazione italiana progettisti sociali APIS

L'Intelligenza artificiale generativa sta trasformando la progettazione sociale, ponendo sfide etiche e metodologiche. Partendo da una definizione teorica del valore sociale, questo contributo analizza criticamente il suo impatto, evidenziando rischi e opportunità. Si esplora il futuro della progettazione sociale e del ruolo del progettista sociale, delineando scenari per un'integrazione responsabile dell'IA che garantisca processi partecipativi e valorizzi il contributo di tutti gli attori coinvolti.

Generative Artificial intelligence is transforming social design, raising both ethical and methodological challenges. Building on a theoretical definition of social value, this contribution critically examines its impact, highlighting both the risks and opportunities it presents. The article explores future directions for social design and the role of the social designer, outlining scenarios for a responsible integration of AI that ensures participatory processes and recognises the value of contributions from all stakeholders involved.

DOI: 10.53223/Sinappsi_2025-02-4

Citazione
Bortolami A., Capone F., Caravello G., Finazzi Agrò A., Maggiora F., Mazzucato A. (2025), La progettazione sociale e la sfida dell'Intelligenza artificiale, *Sinappsi*, XV, n.2, pp.46-61

Parole chiave
Economia sociale
Etica sociale
Intelligenza artificiale

Keywords
Social economy
Social ethics
Artificial intelligence

Introduzione

> *"Per quanto posso ragionevolmente interpretare, tu pensi che secondo me la mia memoria sia l'unico mondo reale e che io affermi che non esiste il mondo esterno… Niente affatto. Nei tuoi termini dovrei essere definito piuttosto un esempio supremo di comunitarianismo oggettivo. Ho in memoria la somma di una storia collettiva, l'insieme di tutte le asserzioni rilevanti fatte dai miei istruttori sul loro mondo esterno, sui loro linguaggi, e sul modo in cui essi usano il linguaggio per produrre le immagini del mondo esterno. […] Io non sono un soggetto, sono la memoria collettiva degli Antipodiani.* ***Non sono un Io, sono un Ciò****. Questo spiega perché posso interagire così bene con ciascuno dei miei istruttori"* (Eco 1999, 317).

Questo contributo si propone di esplorare l'impatto dell'Intelligenza artificiale (IA) generativa sulla progettazione sociale e sul ruolo del progettista sociale, mettendo in luce le opportunità e le sfide emergenti. L'IA generativa, rappresentata da modelli avanzati come i Generative pre-trained transformer (GPT), pre-addestrati su vasti volumi di dati e in

grado di generare autonomamente testi, analizzare dati e individuare schemi ricorrenti, sta assumendo un ruolo sempre più centrale in una pluralità di settori. Questo fenomeno sta trasformando profondamente la nostra realtà, ridefinendo aspetti chiave dell'esistenza umana, quali il lavoro, le relazioni interpersonali e la partecipazione sociale. Grazie ai suoi recenti sviluppi, l'IA generativa ha ampliato notevolmente le sue potenzialità, non solo nel migliorare la produttività e la crescita economica, ma anche nel promuovere il bene sociale (Bankhwal *et. al.* 2024)[1]. In questo contesto di crescente adozione e impiego dell'IA generativa, la progettazione sociale, intesa come il processo di ideazione e sviluppo di interventi volti a migliorare la qualità della vita nelle comunità e rispondere ai bisogni collettivi, appare come un ambito particolarmente suscettibile di trasformazioni significative. Negli ultimi anni, infatti, anche il Terzo settore ha iniziato a esplorare con crescente interesse le opportunità offerte dall'IA generativa, riconoscendola come un prezioso strumento di supporto. Secondo una recente ricerca (Di Troia *et al.* 2024), il 66% degli intervistati del settore non profit dichiara che la propria organizzazione utilizza già qualche forma di IA, mentre il 77% ritiene che un impiego più esteso di queste tecnologie, soprattutto nelle attività legate alla mission, potrebbe apportare benefici significativi.

L'integrazione dell'IA generativa nella progettazione sociale offre numerose opportunità, ma solleva anche questioni critiche che meritano un'attenzione particolare, soprattutto in relazione al ruolo e alle competenze del progettista sociale. Alcuni interrogativi, in particolare, emergono come cruciali: può l'IA realmente comprendere e interpretare in modo adeguato i bisogni sociali? È possibile ridurre dinamiche complesse a semplici dati quantitativi senza rischiare di ignorare aspetti più profondi e qualitativi? Come si può preservare l'intuizione e l'esperienza umana nella progettazione sociale, evitando che l'automazione diventi dominante? Anche le questioni etiche e normative richiedono un approfondimento. L'utilizzo di dati sensibili e la partecipazione di individui vulnerabili comportano responsabilità importanti legate alla privacy, all'equità e alla trasparenza. Quali misure sono necessarie per garantire che i processi siano rispettosi dei diritti individuali e che le soluzioni siano inclusive ed etiche? Come evitare danni derivanti da un uso irresponsabile della tecnologia?

A partire da un inquadramento teorico del valore sociale, che è alla base della progettazione sociale e del ruolo del progettista, il contributo si propone di esplorare, in modo critico, come l'IA possa essere utilizzata in questo ambito. In particolare, verranno analizzati i seguenti aspetti:

- le implicazioni dell'uso dell'IA nella valutazione ex ante dei progetti da parte degli enti finanziatori, considerando l'equilibrio tra misurabilità e valore qualitativo;
- le opportunità e le sfide che l'IA pone al progettista sociale, incluse le possibili trasformazioni nel suo ruolo e nelle sue competenze;
- il rapporto tra IA e comprensione dei bisogni sociali, con un focus sulla capacità dell'IA di cogliere dinamiche complesse e multidimensionali;
- l'equilibrio tra creatività umana e automazione, esplorando come integrare l'IA senza sacrificare l'innovazione e il pensiero strategico;
- le questioni etiche e normative connesse all'uso dell'IA nella progettazione sociale, riguardanti il trattamento dei dati sensibili, la privacy, l'equità e la trasparenza.

In questo modo si vuole contribuire a delineare una visione futura in cui l'IA possa essere integrata nella progettazione sociale come uno strumento complementare, senza compromettere la centralità della partecipazione umana, delle competenze

1 Bankhwal *et al.* (2024) nel report *AI for social good: Improving lives and protecting the planet* riferiscono: "In a recent survey of more than 4,000 not-for-profits conducted by Google for Nonprofits, 75 percent of respondents said that generative AI had the potential to transform their marketing efforts by enhancing their translation and fact-checking capabilities. […] To map the breadth of AÌs applicability, we have developed a database of AI use cases, each of which highlights a type of meaningful problem whose solution could be enabled by one or more AI capabilities. At the time of our 2018 report, this database contained about 170 high-potential use cases. It now contains about 600 – more than a threefold increase. This number is growing as more innovative uses come to light, as social impact leaders continue to experiment boldly, and as AI tools become more accessible and easier to use. […] The experts we surveyed agreed that AI has particularly high potential to make a difference for five SDG goals: Good Health and Well-Being (SDG 3), Quality Education (SDG 4), Affordable and Clean Energy (SDG 7), Sustainable Cities and Communities (SDG 11), and Climate Action (SDG 13). In fact, 60 percent of not-for-profit AI for social good deployments were in these areas" (p. 3).

relazionali e dell'approccio partecipativo, elementi fondamentali per generare un impatto sociale positivo e sostenibile.

1. IA e il valore sociale, tra genesi e valutazione ex ante dei contenuti: presupposti teorici e implicazioni semiotiche

Il valore sociale non è da tempo percepito dalle scienze filosofiche e sociali come qualcosa di fisso. Esiste un ampio consenso sul fatto che il valore sociale non sia un dato statico, ma un'entità culturale soggetta a ridefinizione e trasformazioni continue. In quanto valore collettivamente apprezzato, non è colto come un'entità ipostatica, né come un dato obiettivo e cristallizzato esterno alla comunità che lo esprime. È invece percepito come il risultato di un complesso processo di stipule e contro stipule sociali, di negoziazioni collettive, cioè in quanto meccanismo di tipo storico culturale, attraverso cui individui e gruppi definiscono, contrattano e istituzionalizzano ciò che viene considerato significativo all'interno di una comunità. Questo processo si fonda su dinamiche di interazione, in cui il riconoscimento reciproco e le convenzioni condivise attribuiscono valore a determinati oggetti, comportamenti o norme. Come suggerito da Berger e Luckmann (2016) nella loro teoria della costruzione sociale della realtà, il valore emerge inizialmente da pratiche sociali reiterate, che nel tempo si consolidano in istituzioni capaci di perpetuare e legittimare tali significati.

Nel contesto italiano, anche in ambienti per tradizione più attenti a un approccio fondazionista di una possibile assiologia universale, non mancano voci critiche, come quelle che nel dibattito cattolico sui cosiddetti 'valori non negoziabili' hanno problematizzato il ricorso a tale espressione, facendo valere la sua scarsa attestazione nella tradizione e nel più recente Magistero sociale della Chiesa[2]. Analogamente, Chiodi *et al.* (2012) affermano: "se il riferimento alla dignità della persona è ciò che rende sensato il linguaggio dei valori, tale dignità può a sua volta essere declinata solo praticamente, nel confronto con le mutevoli e complesse circostanze della vita umana" (p. 664).

Il valore sociale, come risultato di un processo negoziale e collettivo, è anche lo specifico oggetto della valutazione delle proposte progettuali, che deve tenere conto delle dinamiche sociali e culturali che lo costituiscono. In questo contesto, la valutazione ex ante nei processi amministrativi che regolano il Ciclo di vita di Progetto (Formez 2002) rappresenta un momento interpretativo cruciale che precede il finanziamento pubblico da parte del titolare del procedimento (stazione appaltante, committente ecc.). La Commissione valutatrice, composta da membri interni alla Pubblica amministrazione e agli Enti erogatori privati o da membri interni ed esperti esterni, ha lo scopo di determinare il valore sociale che una proposta progettuale può scatenare in un dato contesto. Al fine di garantire obiettività e affidabilità all'intero procedimento e prevenire possibili arbitri amministrativi, è prassi consolidata che, già in fase di pubblicazione dell'avviso o del bando, vengano definiti specifici 'criteri di valutazione', trasparenti e dettagliati. Tali criteri sono generalmente articolati in scale di punteggio relative alle singole dimensioni o aree di valutazione, che a loro volta possono dischiudere un campo discrezionale più o meno ampio in capo al soggetto valutatore. Sebbene sia consentita una certa latitudine di giudizio al valutatore, questa deve sempre mantenersi entro i confini stabiliti dai criteri, per evitare derive arbitrarie e garantire la trasparenza e obiettività dell'azione amministrativa.

Una tendenza recente è tuttavia ridurre progressivamente tale margine di discrezionalità, fino a eliminarla del tutto, attraverso procedure di attribuzione del punteggio completamente

2 L'espressione 'valori non negoziabili' si è attestata nel dibattito politico e sociale a partire dal 2006, in seguito al discorso di Benedetto XVI ai parlamentari del Partito popolare europeo. Pur non letteralmente riportata nel testo del discorso – dove il Pontefice si riferì a "principi che non sono negoziabili" – l'espressione si è diffusa rapidamente, annoverando tra i suoi sostenitori anche figure esterne al mondo cattolico. Per 'valori non negoziabili' si è inteso quel nocciolo di principi morali e posizioni etiche, riguardanti sia la vita sociale sia le opzioni personali, sottratti al confronto e astrattamente non contendibili nel dibattito pluralistico perché fondati sulla 'natura' (cioè, la *lex naturalis*) e sulla 'ragione universale'. Giova ricordare che, benché il pensiero giusnaturalistico, specie quello di tradizione scolastica, rappresenti un riferimento costante per l'evoluzione della dottrina morale e sociale della Chiesa cattolica, in questi termini, in tutti gli atti magisteriali recenti e meno recenti della Santa Sede, e anche nei documenti ufficiali della Conferenza episcopale italiana (CEI), l'espressione non ricorre mai. In Italia, invece, è stato costante, nella fase conclusiva del pontificato di Giovanni Paolo II e di Benedetto XVI poi – così come durante le fasi delle presidenze CEI di Camillo Ruini (1991-2007) e Angelo Bagnasco (2007-2017) – il richiamo a principi metastorici, dati per ovvi e autoevidenti, specialmente in materia di bioetica, morale sessuale e familiare.

oggettive e automatizzate. Un esempio di questa evoluzione è rappresentato dall'Avviso n. 2/2024 del Ministero del Lavoro e delle politiche sociali, *per il finanziamento di iniziative e progetti di rilevanza nazionale ai sensi dell'articolo 72 del decreto legislativo del 3 luglio 2017, n. 117 – Anno 2024*[3]. Non c'è dubbio che l'abbattimento dei tempi amministrativi, attraverso l'attestazione automatica dei requisiti e, in futuro, l'uso di IA per l'interpretazione testuale e l'attribuzione dei punteggi, potrebbe portare vantaggi in termini di risparmio, obiettività e controllo multi-stakeholder. Tuttavia, un uso eccessivo dell'automazione e la completa eliminazione della discrezionalità umana basata sulla competenza nella valutazione delle proposte comportano rischi significativi. Un tale approccio potrebbe alimentare l'impersonalità e l'automatismo nell'azione amministrativa, rischiando di accentuare la deriva 'tecnocratica' della burocrazia, che è tra i rischi paventati da Max Weber (2001). La burocrazia, pur essendo una delle espressioni più avanzate della razionalizzazione del mondo moderno, potrebbe finire per diventare una 'gabbia d'acciaio' (*eine stahlhartes Gehäuse*), in cui gli individui rischiano di perdere libertà e umanità, schiacciati da regole e strutture.

La questione centrale è intendersi su che tipo di processo rappresenti la progettazione sociale e, in subordine, quali siano i suoi principali output e outcome nella catena del valore. Secondo la recente Norma UNI del 30 aprile 2019 n. 11746, *Attività professionali non regolamentate - Progettista Sociale - Requisiti di conoscenza, abilità e competenza*, il progetto sociale è definito come "un intervento sistemico e organizzato, partecipato da più soggetti, orientato a produrre risultati apprezzabili e direttamente finalizzato all'interesse generale"[4]. Esso offre un valore aggiunto alla società, in quanto effettua la ricognizione, l'attivazione e la messa in sinergia di risorse, capacità e disponibilità potenzialmente presenti, per sviluppare iniziative civiche, solidaristiche e di utilità sociale, al fine di perseguire l'interesse generale.

La progettazione sociale non può essere ridotta a una semplice procedura burocratica chiusa e circoscritta alla diade committente/proponente che sfocia in un prodotto documentale di tipo adempitivo, facilmente emulabile tramite procedure automatizzate rese possibili dall'attuale sviluppo dell'IA generativa. Al contrario, la progettazione sociale e gli output documentali che le appartengono sono il frutto di un processo decisionale collettivo, quindi un processo 'politico', come già indicato nello standard del Project Life Cycle Management (Formez 2002). Le fonti autentiche di tale processo non sono dataset di carattere enciclopedico, statistico, socioeconomico, cioè essenzialmente un sistema semiotico di rinvii interni di espressioni e loro possibili 'interpretanti' distribuiti in tutte le possibili selezioni contestuali, 'enciclopedie', nei termini di Eco (1999), interconnesse in una memoria globale' (pp. 304-324) – molto simile a ciò che oggi l'IA è in grado di simulare – su cui può efficacemente attestarsi un sistema di generazione basato sul calcolo statistico combinatorio.

Regesti di tale genere possono certamente sostenere un processo decisionale di tipo socio-politico, ma le sue radici andranno sempre ricercate in mozioni, aspettative, interessi, obiettivi collettivi o gruppali che, come ancora osservano Berger e Luckmann (2016) rifacendosi a Edmund Husserl (2008), promanano da 'mondi di vita' (*Lebenswelt*) irriducibili alla tecnica e alla razionalità strumentale, e quindi irriducibili a qualunque sistema semiotico, inclusi i sistemi di creazione e processamento testuale generativi messi a disposizione dall'IA. In definitiva, anche il progetto sociale, in quanto prodotto testuale, deve soddisfare le "regole conversazionali" avanzate da Grice (1989): (i) regola della Quantità: fornisci tutta l'informazione necessaria; (ii) regola della Qualità: non dire ciò che credi falso. Non dire ciò per cui non hai prove adeguate; (iii) regola della

3 L'Avviso in argomento ha adottato criteri di valutazione e selezione delle proposte, esplicitati al § 11. *Concessione del contributo*, interamente basati su elementi obiettivi e documentabili, su base curriculare o di autodichiarazione, da parte dei proponenti. Il punteggio totalizzabile, quindi, è esclusivamente funzione della sommatoria dei requisiti relativi a: esperienza, ampiezza e struttura della rete della proposta, distribuzione territoriale e apporto finanziario. Tali elementi sono qualificati dai proponenti attraverso una tabella Excel, da compilarsi tra gli allegati all'Avviso che, sebbene rudimentalmente, consente il calcolo automatico del punteggio finale totalizzato (fino ad un massimo di 80 punti). Questo sistema permette, in prospettiva, la formazione delle graduatorie di merito finali per il finanziamento delle proposte.

4 UNI (2019), Norma n. 11746 *Attività professionali non regolamentate - Progettista sociale - Requisiti di conoscenza, abilità e competenza*, 30 aprile 2019.

Relazione (o Pertinenza): sii pertinente; (iv) regola del Modo: evita oscurità di espressione. Se i diversi dispositivi di IA, sempre più avanzati, possono soddisfare le regole (i), (iii) e in parte (iv), è assai dubbio che al di fuori del controllo umano possano soddisfare la (ii), in quanto l'attendibilità cui la regola richiama si basa su un presupposto estensionale ed extra semiotico, e non solo intensionale, cioè interno al piano del contenuto culturale già depositato che costituisce, attraverso diverse forme di prelievo, il 'thesaurus' dell'IA (p. 26).

Il rischio anzitutto etico e politico, e poi prasseologico e metodologico, è pertanto che un ricorso scarsamente consapevole ai sistemi di valutazione basati sull'IA nella valutazione dei progetti sociali, come pure nella loro produzione, finisca per simulare la vita sociale cristallizzandola in un sistema chiuso, laddove invece i sistemi sociali sono per definizione *lebenswelt*, aperti, interattivi e iterativi.

2. Il ruolo del progettista sociale nell'era dell'IA

La progettazione sociale è un processo metodologico volto a identificare, pianificare e realizzare interventi che mirano a migliorare le condizioni di vita e il benessere delle persone e delle comunità. Questo tipo di progettazione si concentra su problematiche sociali specifiche, come povertà, esclusione sociale, disuguaglianze, accesso ai servizi, educazione, salute e sviluppo comunitario, con l'obiettivo di rispondere in modo concreto e sostenibile ai bisogni delle persone. A differenza di altre forme di progettazione, la progettazione sociale richiede competenze umane come ascolto, negoziazione e mediazione, essenziali per costruire fiducia, facilitare la partecipazione e trovare soluzioni adeguate ai contesti sociali. È un ambito antropocentrico, centrato sull'uomo in tutte le fasi del processo, dalla comprensione del problema alla creazione e implementazione della soluzione (Norman 2019). Conseguentemente, il progettista sociale "nella propria attività promuove processi operativi basati sull'integrazione di soggetti pubblici, privati e gruppi informali, lo scambio tra livelli decisionali di indirizzo politico, attuativo, operativo, l'attivazione di reti e di partenariati strutturati tra organizzazioni diverse, la collaborazione tra varie professionalità, la partecipazione dei cittadini attivi, dei destinatari degli interventi, delle persone motivate all'interesse collettivo. Svolge la propria funzione prevalentemente come mediatore tra soggetti finanziatori (che esprimono priorità, procedure, regolamentazioni, criteri vincolanti), soggetti attuatori (con le loro mission, competenze, legami territoriali, risorse, identità) e contesto sociale (con domande sociali, dinamiche relazionali, reti formali e informali, sistema di servizi, cultura locale)"[5]. Il progettista sociale è il professionista che guida il processo di progettazione, dalla fase di analisi dei bisogni alla realizzazione e monitoraggio degli interventi. Il suo ruolo va oltre quello di un semplice tecnico, poiché è chiamato a svolgere diverse funzioni, tra cui:

Analisi dei bisogni: il progettista sociale deve raccogliere e interpretare dati sul contesto sociale, economico e culturale della comunità, al fine di identificare i bisogni espliciti e latenti. Questo implica un ascolto attivo e una comprensione profonda della realtà sociale, spesso attraverso interviste, focus group e altri strumenti di partecipazione.

Progettazione e *pianificazione*: dopo aver compreso i bisogni, il progettista sociale deve elaborare interventi che possano rispondere efficacemente a tali bisogni. Questo processo implica la definizione di obiettivi, strategie e azioni specifiche, nonché la pianificazione delle risorse necessarie. Il progettista sociale deve anche considerare la sostenibilità degli interventi, prevedendo come questi potranno essere mantenuti nel lungo periodo.

Gestione e coordinamento: il progettista sociale ha la responsabilità di coordinare tutte le fasi del progetto, garantendo che le risorse siano utilizzate in modo efficiente e che gli obiettivi siano raggiunti. Ciò comporta la gestione di gruppi di lavoro, la comunicazione con i diversi stakeholder e la garanzia che il progetto rimanga aderente ai principi etici e alle normative di riferimento.

Facilitazione della partecipazione: un aspetto cruciale della progettazione sociale è la partecipazione delle persone e delle comunità. Il progettista sociale deve lavorare come facilitatore, creando spazi di dialogo e di condivisione in cui tutti gli attori coinvolti possano contribuire attivamente alla definizione e attuazione del progetto.

5 UNI (2019), Norma n. 11746 *Attività professionali non regolamentate - Progettista sociale - Requisiti di conoscenza, abilità e competenza*, 30 aprile 2019.

Monitoraggio e valutazione: una volta realizzato il progetto, il progettista sociale è chiamato a monitorare l'efficacia dell'intervento e a valutarne gli impatti. Questo processo implica la raccolta di dati sul campo, l'analisi dei risultati ottenuti e l'adattamento delle azioni, se necessario, per migliorare l'efficacia del progetto nel tempo.
Gestione del cambiamento: il progettista sociale deve essere in grado di gestire il cambiamento sociale attraverso il coinvolgimento delle persone, delle istituzioni e degli altri attori, facilitando la transizione verso soluzioni più inclusive e sostenibili. Deve essere un mediatore tra le diverse istanze e lavorare per conciliare gli interessi e le esigenze delle varie parti coinvolte.

L'uso dell'IA generativa sta modificando queste dinamiche, da una parte influenzando il modo in cui si individuano i bisogni, si costruiscono le ipotesi di intervento e si valutano le priorità, dall'altra sollevando dubbi sulla possibile perdita di controllo umano e sul rischio di allontanamento dai reali bisogni delle comunità. Un esempio emerso da un confronto tra progettisti sociali evidenzia questa tensione: in un Comune rurale, l'IA generativa ha analizzato dati epidemiologici e socioeconomici, individuando un urgente bisogno di accesso ai servizi medici per la popolazione anziana, distante dai centri sanitari. Tuttavia, il progettista ha notato che l'analisi automatizzata trascurava altri bisogni rilevanti come quelli delle famiglie monoparentali a basso reddito. Grazie a incontri partecipativi e questionari, l'analisi è stata integrata con i bisogni sociali emersi, portando a soluzioni mirate come trasporti agevolati per anziani e un fondo per famiglie vulnerabili.

Questo caso mostra come l'IA stia diventando utile nella fase analitica, ma anche come l'automazione possa lasciare dei vuoti che richiedono un intervento umano per essere colmati. Tale aspetto trova conferma in una ricerca sull'impatto dell'IA nel project management, secondo cui, il 73% degli intervistati ritiene che l'IA non potrà mai sostituire del tutto l'intelligenza e la sensibilità umana (Alshaikhi e Khayyat 2021). Proprio per questo, sebbene l'IA generativa possa supportare molte fasi del ciclo di vita del progetto, come la scrittura, la raccolta di dati e la reportistica, la sua efficacia dipende dalla capacità del progettista di istruirla, interrogarla, interpretarne i dati e contestualizzarli.

Senza un'adeguata supervisione, l'IA rischia di generare proposte standardizzate, prive di una reale comprensione delle dinamiche sociali (Mazzucato *et al.* 2025; Barcaui e Monat 2023; Alshaikhi e Khayyat 2021). Di conseguenza, il ruolo del progettista sociale si trasforma: non è più solo autore e gestore di progetti, ma anche supervisore e facilitatore dell'IA. Le sue competenze si concentrano sulla definizione degli input per i modelli di IA e sulla revisione critica degli output generati. Un esempio concreto è dato da un progetto di inclusione sociale per migranti, in cui il progettista ha utilizzato l'IA per analizzare i bisogni emergenti e creare un piano d'azione. Ha fornito input specifici all'algoritmo, includendo dati qualitativi raccolti tramite focus group con i beneficiari. Quando l'IA ha suggerito priorità basate esclusivamente su una visione economica del problema, il progettista ha rivisto gli output, integrando aspetti umani come il supporto psicologico e l'accesso alla formazione culturale, trascurati dai dati numerici.

L'evoluzione professionale si riflette anche nell'automazione di alcune attività, che permette di liberare tempo prezioso per potenziare gli aspetti relazionali e strategici, garantendo così un impatto più significativo nelle aree in cui l'intervento umano è imprescindibile. L'IA generativa, in tal senso, può rappresentare un 'elemento accelerante' dell'evoluzione di una professione spesso 'schiacciata' dalla burocrazia progettuale imposta dai finanziatori, che finisce per limitare la funzione di garante della strategia dell'intervento. Essa rappresenta senza dubbio un 'supporto intelligente' per i progettisti, in grado di rendere più efficienti i processi di stesura e rendicontazione, riducendo l'impegno nelle attività gestionali. Tuttavia, è essenziale che il suo utilizzo non comprometta la dimensione valoriale della progettazione sociale, che rimane un processo fondato su principi di partecipazione, inclusione e coesione sociale. Come ogni tecnologia, l'IA generativa è uno strumento abilitante, non sostitutivo del progettista. Nel sociale, l'attività di progettazione è il momento privilegiato in cui si manifestano i propri valori, i modelli di riferimento e le teorie del cambiamento che vengono utilizzati, spesso in modo implicito e talvolta inconsapevole, per dare significato e spiegare i fenomeni sociali. In questo processo, si confrontano anche diversi sistemi di valori, che influenzano le scelte e le soluzioni adottate (Cascino 2024).

'Progettare' (dal latino *proicere*, 'gettare avanti') implica anticipare l'azione e il processo, ossia le sequenze di atti che dovrebbero condurre a un risultato di cambiamento precedentemente definito (Sicora e Pignatti 2024). Nella progettazione sociale, la dimensione ideativa è fondamentale, ma è altrettanto importante considerare come vengono individuati i bisogni sociali o intercettati i problemi latenti, e come questi vengono risolti o soddisfatti. La progettazione sociale non si basa su dati astratti, ma agisce in contesti reali e complessi, dove si intrecciano dinamiche culturali, economiche e politiche. In questo scenario in evoluzione, il progettista sociale non si limita a supervisionare l'uso dell'IA, ma mantiene un ruolo centrale come interprete delle dinamiche sociali e garante delle risposte ai bisogni delle comunità. Per quanto 'la macchina' possa agevolare l'analisi di grandi quantità di dati e di informazioni utili, difficilmente potrà sostituire il progettista sociale e la sua capacità interpretativa e predittiva, di comprensione e di decodifica di fenomeni e contesti complessi e dinamici, nonché dei trend sui quali intervenire. L'IA, piuttosto, potrà aiutare il progettista a 'non lanciare il cuore oltre l'ostacolo' e a 'non superare i limiti' della fattibilità degli interventi, consentendogli di migliorare l'efficacia di questi ultimi e accrescere il loro valore generato. L'IA potrà individuare la 'strada giusta' ma difficilmente riuscirà a disegnarla.

Nonostante la tendenza attuale verso la standardizzazione dei processi e la semplificazione dei fenomeni sociali, il progettista sociale diventa un attore chiave nel definire un utilizzo responsabile e inclusivo dell'IA. Come tale, deve adattarsi ai cambiamenti in corso e aprirsi a nuove competenze per affrontare le sfide future, assicurando che la progettazione sociale resti ancorata ai valori e alle necessità delle comunità. Per integrare strumenti di supporto come l'IA, il progettista sociale deve sviluppare competenze specifiche, tra cui il prompting, l'analisi critica, il problem solving e la padronanza degli strumenti digitali. Deve essere in grado di 'interrogare' l'algoritmo, garantendo decisioni informate ed evitando i *bias* algoritmici che potrebbero compromettere l'equità degli interventi. Come afferma Norman (2019), la progettazione sociale non è una mera procedura tecnica, bensì un processo umano che richiede sensibilità alle dinamiche sociali. Nell'ambito di un progetto per ridurre la disoccupazione giovanile, ad esempio, un progettista ha usato un algoritmo di IA per abbinare giovani e opportunità formative e lavorative. Ha sviluppato competenze di prompting per porre domande specifiche all'algoritmo, come l'incrocio tra competenze digitali dei candidati e domanda del mercato locale. Analizzando i risultati, ha individuato un *bias* che escludeva i candidati delle aree periferiche, a causa della scarsa connettività evidenziata dai dati. L'intervento umano ha corretto la distorsione, garantendo un approccio più equo e inclusivo.

La 'revisione' del bagaglio professionale del progettista sociale, quindi, diventa un passaggio obbligato: il progettista deve guardarsi dentro, essere consapevole del proprio ruolo sociale e delle innumerevoli competenze a disposizione per fronteggiare la convivenza con l'IA generativa. È importante, inoltre, che alimenti la sua curiosità e l'entusiasmo nell'esplorare le 'vaste praterie' delle opportunità offerte dalle nuove tecnologie.

L'IA non dovrebbe sopraffare il progettista, ma rafforzarlo, consentendogli di sviluppare competenze inesplorate e valorizzare abilità 'assopite', così da distinguersi dalla 'macchina'. L'IA può offrirgli l'opportunità di riscoprire la bellezza della professione, il suo valore strategico per la comunità, l'importanza delle relazioni e del confronto, così come le sue capacità interpretative e di lettura dei contesti sociali e delle loro complessità. Alleggerendo i compiti più amministrativi, permette al progettista di concentrarsi maggiormente sull'evoluzione professionale nonché sul miglioramento e potenziamento delle proprie ritrovate e innovate competenze, sempre più orientate al bene comune e alla generazione di valore sociale (Abbass 2019).

3. IA, creatività e comprensione dei contesti sociali

Per approfondire l'impatto dell'IA su aspetti fondamentali del lavoro del progettista sociale, quali la comprensione dei contesti sociali e il coinvolgimento degli attori sociali in un'ottica partecipativa, è necessaria una breve premessa relativa ai differenti modelli teorici e approcci alla progettazione (Leone e Prezza 2003). Secondo Arnstein (1969), i processi di progettazione sociale possono essere analizzati attraverso la 'scala della partecipazione', enfatizzando il livello di coinvolgimento. In base al grado di pre-strutturazione e apertura, possiamo identificare tre tipologie di approcci e orientamenti nella progettazione di interventi nel sociale:

Approccio sinottico-razionale. Partendo da ipotesi di causalità lineare tra le problematiche sociali, il progetto viene costruito a partire da una

comprensione a priori dei bisogni, estranea e lontana dai soggetti 'portatori' del problema. Il processo di progettazione rimane una competenza specifica del progettista, il cui compito è quello di trovare i mezzi migliori per raggiungere obiettivi dati, in condizioni ambientali statiche. Nonostante diversi autori abbiano messo in discussione l'idea di una razionalità assoluta degli individui (Kahneman e Tversky 1979; Simon 1947; 1985), questo modello meccanicista, poco adatto a contesti complessi come quelli sociali, è ancora prevalente in molti contesti progettuali e nei sistemi di valutazione di bandi pubblici e privati. In tal senso, Innes e Booher (2010) evidenziano la sua inadeguatezza nel trattare problemi sociali caratterizzati da ambiguità, conflitti di valore e instabilità, proponendo un passaggio verso forme di razionalità collaborativa. Anche Siza (2018) sottolinea come i nuovi contesti richiedano maggiore flessibilità e apertura, superando l'illusione di poter predeterminare soluzioni ottimali.

Approccio concertativo o partecipativo. In questo approccio di impronta costruttivista, si ritiene che i problemi sociali non seguano una causalità lineare, ma siano caratterizzati da molteplici letture dei bisogni e delle ipotesi interpretative. Il ruolo del progettista non è quello di distribuire soluzioni, ma di promuovere empowerment a livello individuale e di comunità. L'interazione tra i diversi attori impegnati è fondamentale e caratterizza tutte le tappe del percorso di costruzione di un intervento sociale. In tale contesto si inserisce il contributo di Ledwith (2016), il quale, facendo esplicito riferimento al pensiero di Freire, propone un modello di sviluppo comunitario basato sulla formazione di coscienza critica, sull'azione collettiva e sul protagonismo dei cittadini. L'approccio partecipativo contemporaneo si configura quindi come uno spazio di apprendimento condiviso e trasformativo, piuttosto che come semplice consultazione.

Approccio euristico. Questo approccio sostiene che focalizzarsi troppo sul 'prodotto' finale può distogliere l'attenzione dai processi e dai risultati reali di un intervento sociale. L'approccio euristico pone al centro l'attivazione, da cui emergono molteplici sotto-progetti, tra loro connessi, ciascuno con il proprio percorso di progettazione, realizzazione e verifica. A differenza del modello sinottico-razionale, il progetto non è definito a priori. In questa prospettiva risulta utile il contributo di Manzini (2015) che descrive la progettazione sociale come un processo emergente e collaborativo, in cui il progettista non impone soluzioni, ma facilita percorsi di co-creazione in cui anche i cittadini assumono un ruolo attivo e creativo. L'approccio euristico, come evidenzia lo stesso autore, privilegia la sperimentazione e la riflessività, riconoscendo che le soluzioni si costruiscono nel dialogo con il contesto e non tramite l'applicazione di modelli rigidi. Il passaggio dall'approccio sinottico a quello euristico nella progettazione sociale evidenzia l'importanza crescente delle dinamiche relazionali e della comprensione profonda dei contesti sociali.

In questo contesto, la domanda cruciale riguarda l'efficacia dell'IA nel comprendere aspetti come bisogni psicologici, dinamiche non esplicite ed emozioni, che sono fondamentali per una progettazione sociale valida.

Studi recenti (Benasayag e Pennisi 2024; Wang *et al.* 2024) si sono focalizzati su due aspetti: l'intelligenza sociale, intesa come l'abilità di agire in maniera appropriata in situazioni che implicano gli scambi relazionali, e la capacità dell'IA di comprendere comportamenti complessi. Essi hanno evidenziato che i modelli di IA, come GPT, mostrano una comprensione basilare dell'intelligenza sociale, mentre gli esseri umani operano a livelli significativamente più avanzati.

Un esempio utile per comprendere questi limiti riguarda il riutilizzo a fini sociali di un bene confiscato alla criminalità organizzata. L'elaborazione dell'intervento non si è basata solo su analisi oggettive dei bisogni o su dati socioeconomici del territorio. È risultato determinante comprendere la percezione e il vissuto sedimentato intorno a quel luogo: per molti cittadini rappresentava un simbolo di paura, omertà e collusione. L'IA, pur potendo contribuire all'analisi di dati relativi alla marginalità o alla disoccupazione giovanile, non sarebbe stata in grado di intercettare il peso emotivo e simbolico che quel luogo incarnava per la comunità. Solo attraverso una presenza prolungata sul territorio, incontri pubblici, laboratori partecipativi e pratiche di ascolto attivo è stato possibile far emergere questi aspetti ed elaborare un progetto in grado di attivare un 'processo di riconquista simbolica' del bene. Alla luce di questa esperienza, ciò che viene percepita come una capacità dell'IA di cogliere sfumature sociali è, in realtà, il risultato della sua abilità di individuare pattern ricorrenti basandosi su correlazioni statistiche, senza una reale capacità

di comprensione empatica o esperienziale[6]. Non si tratterebbe, quindi, di una vera intelligenza sociale paragonabile a quella umana quanto di una simulazione, seppur estremamente accurata. Questo implica che l'IA può svolgere una funzione primaria nella progettazione sociale solo all'interno di un approccio sinottico-razionale. Al contrario, negli approcci partecipativi ed euristici, che richiedono *soft skills* come empatia, lavoro di squadra e capacità comunicative, relazionali e di negoziazione, l'IA presenta evidenti limitazioni. Tali approcci enfatizzano l'importanza dell'interazione umana nella costruzione di fiducia tra gli attori sociali, nella costruzione di partenariati, nell'attivazione di processi partecipativi e nella mediazione dei conflitti, attività che l'IA non è in grado di replicare.

Ciascun contesto sociale è caratterizzato da un complesso di elementi culturali, ideologici, sociologici, economici che lo rendono peculiare, e uniche sono le strategie che l'intervento progettuale mette in atto per rispondere alle esigenze specifiche. In questo scenario gioca un ruolo centrale la creatività: la capacità di generare idee, soluzioni o concetti nuovi e originali, combinando in modo innovativo elementi preesistenti. È un processo che permette di andare oltre le soluzioni standardizzate, immaginando nuove opportunità che rispondano alle specificità di ogni contesto e costruendo soluzioni innovative che migliorano le condizioni di vita, rafforzano le relazioni e promuovono il cambiamento. Come evidenziato da Smith e Gün (2004), la capacità di riconoscere connessioni inedite, interpretare bisogni non espliciti e costruire scenari futuri è un elemento distintivo dell'intervento umano nei contesti sociali. Per questo motivo la progettazione sociale mette in campo competenze che vanno ben oltre le capacità tecniche. Attraverso la creatività, i progettisti sociali assumono un ruolo cruciale nel tradurre bisogni complessi in soluzioni pratiche, che siano in grado di favorire la coesione sociale, stimolare la partecipazione collettiva, innescare processi trasformativi non solo in termini di condizioni materiali, ma anche di relazioni, di esperienze e dinamiche culturali. Questo implica la necessità di ascoltare i bisogni espressi e, al contempo, cogliere le sfumature relazionali e culturali che influenzano le dinamiche del contesto sociale. Il progettista sociale non è solo un tecnico, ma anche un facilitatore di dialoghi e processi che promuovono il cambiamento. Tale approccio si basa sull'empatia, sull'ascolto e sulla capacità di trasporre intuizioni complesse in azioni concrete. L'empatia incoraggia i progettisti a valutare e adattare il proprio mindset e le proprie azioni durante tutto il processo di progettazione (Mert 2024). È il caso di un progetto di rigenerazione urbana che, in un quartiere periferico, ha trasformato un edificio abbandonato in un centro di incontro per giovani e anziani, grazie alle capacità di empatia e ascolto dei progettisti, che hanno colto il valore simbolico dello spazio per la comunità. Coinvolgendo i residenti in workshop creativi, sono stati sviluppati usi condivisi dello spazio, come corsi, laboratori e attività ricreative, stimolando così un senso di comunità.

L'IA si presenta come uno strumento promettente per supportare i processi progettuali, ma è intrinsecamente limitata dalla sua natura algoritmica, soprattutto quando è necessario comprendere le specificità del contesto e sviluppare strategie non standardizzate. Sebbene i sistemi di IA siano altamente efficienti nell'elaborazione dei dati e nell'individuazione di pattern, incontrano difficoltà nel cogliere la complessità dei contesti umani, che includono emozioni, dinamiche informali e culturali (Glikson e Woolley 2020). Ad esempio, un algoritmo può identificare correlazioni tra indicatori sociali, ma non è in grado di interpretare le motivazioni sottostanti o le dinamiche che influenzano quei dati, il che può portare a soluzioni che mancano di sensibilità culturale o che rinforzano *bias* preesistenti (Kamila e Jasrotia 2025). Inoltre, la mancanza di una capacità di contestualizzazione rende l'IA inefficace nel rispondere a situazioni complesse e imprevedibili, caratteristiche tipiche dei sistemi sociali (Ryan e Stahl 2020). Un ulteriore limite riguarda l'incapacità dell'IA di stimolare innovazione

6 Questa mancanza compromette l'efficacia in situazioni in cui il linguaggio del corpo e il non verbale sono elementi determinanti. Occorre, infatti, considerare l'aspetto dell'*embodiment* (Plebe 2019). L'IA, pur essendo in grado di raggiungere risultati fino a ieri impensabili, si scontra con il limite della mancanza di una fisicità. Le nostre conversazioni, le nostre espressioni linguistiche e, in generale, le nostre relazioni sono radicate nella corporeità, senza la quale la capacità di comprensione di tali aspetti è fortemente ridotta (Guido 2024 - Intervista alla Prof.ssa Carrozza).
Molti autori sostengono che con lo sviluppo della robotica e del *deep learning* tale limite sarà superato; tuttavia, al momento lo sviluppo dell'*Embodied AI* è ancora in una fase embrionale (Paolo *et al.* 2024).

autentica in contesti dove il cambiamento deve emergere dal basso, attraverso processi partecipativi e intrinseci (Banavar 2016).

La sinergia tra la creatività umana e il potenziale analitico dell'IA può, tuttavia, aprire nuove prospettive per lo sviluppo di soluzioni innovative e sostenibili, migliorando l'efficienza dei processi progettuali senza sacrificare la profondità dell'analisi umana. Come è stato evidenziato nei precedenti paragrafi, l'IA può facilitare e accelerare i processi di analisi dei dati che descrivono il contesto sociale, fornendo ai progettisti sociali una base di informazioni da cui partire. Ad esempio, strumenti di elaborazione del linguaggio naturale possono analizzare grandi quantità di dati qualitativi, identificando temi e bisogni ricorrenti. Le tecnologie di Intelligenza artificiale possono offrire spunti iniziali, liberando tempo per attività più creative e strategiche. Un altro ambito in cui l'IA si configura come uno strumento utile è la modellazione di scenari. Attraverso simulazioni e analisi predittive, l'IA può valutare l'impatto di diverse strategie progettuali in anticipo, fornendo ai progettisti una visione più chiara dei rischi e delle opportunità. Glikson e Woolley (2020) evidenziano che questa capacità può ridurre l'incertezza e migliorare la pianificazione, pur mantenendo il controllo umano sui processi decisionali. Kamila e Jasrotia (2025) aggiungono che questo approccio stimola la creatività senza sostituirla, offrendo nuovi punti di partenza per affrontare sfide complesse.

L'integrazione tra IA e creatività umana richiede dunque un approccio bilanciato, in cui l'IA si prefigura come supporto che valorizza le competenze distintive della professionalità che il progettista sociale mette in campo. La tecnologia può senz'altro aumentare l'efficacia dell'analisi, ma le intuizioni umane rimangono essenziali per garantire che le soluzioni siano rilevanti, etiche e sostenibili, rispettando le specificità del contesto e le necessità dei destinatari dell'intervento progettuale. Glikson e Woolley (2020) propongono che l'IA possa supportare efficacemente le attività analitiche e operative, pur mantenendo il progettista al controllo delle decisioni creative e strategiche. Questo approccio, in cui l'IA funge da strumento complementare, potrebbe ottimizzare i risultati progettuali, promuovendo anche un processo più partecipativo e inclusivo, in linea con i principi della progettazione sociale. Allo stesso modo, l'adattabilità dei sistemi IA, come osservato da Booyse e Scheepers (2023), è fondamentale per affrontare le complessità del mondo contemporaneo, garantendo che l'IA non sostituisca, ma piuttosto completi, l'intuizione e la creatività umana, per affrontare sfide complesse senza mai perdere di vista i valori fondamentali della progettazione sociale: inclusione, partecipazione e trasformazione.

4. Questioni etiche e normative

Studi recenti (Bankhwal *et al.* 2024; Qian *et al.* 2024; Gînguță *et al.* 2023; Zhuo *et al.* 2023) confermano che l'IA, pur offrendo opportunità significative, presenti anche rischi quali *bias* algoritmici, discriminazione, mancanza di trasparenza e violazione della privacy. Questi fattori, come illustrato di seguito, possono incidere sulla qualità dei progetti sociali e sulla vita dei beneficiari, sollevando interrogativi di natura etica e normativa.

Il *bias* algoritmico rappresenta una delle principali criticità dell'Intelligenza artificiale generativa ed è tra i temi più dibattuti negli ultimi anni. Deriva da assunzioni errate nel processo di apprendimento automatico, portando un algoritmo di IA a produrre risultati sistematicamente distorti, con effetti discriminatori e ingiusti nei confronti di gruppi vulnerabili. Le distorsioni possono originarsi dai dati di addestramento, se contengono pregiudizi preesistenti, o dai prompt degli utenti, che possono incorporare stereotipi inconsapevoli. In questo modo, l'IA rischia di amplificare forme di discriminazione basate su fattori come etnia, religione, genere o orientamento sessuale. Quando l'IA produce risposte influenzate dai *bias*, può condizionare decisioni cruciali in ambiti sociali come l'occupazione, la giustizia, l'istruzione, l'accesso ai servizi pubblici, penalizzando specifici gruppi e compromettendo i principi di equità e giustizia sociale (FRA 2022; Zajko 2022; Brynjolfsson e McAfee 2017). Nel contesto della progettazione sociale, un sistema di IA utilizzato per distribuire risorse nell'ambito di un intervento sociale, ad esempio, potrebbe inconsapevolmente svantaggiare alcuni gruppi se i dati su cui si basa riflettono disuguaglianze economiche o sociali preesistenti.

Un concetto strettamente collegato al *bias* algoritmico, di rilievo per chi si occupa di progettazione sociale è quello di "colonialismo dell'IA", introdotto da Mohamed *et al.* (2020). Secondo gli autori, i sistemi di IA non sono globali

e neutrali come spesso si tende a credere, ma riflettono i modelli culturali dei Paesi che li sviluppano e addestrano, ovvero quelli che storicamente occupano posizioni di leadership. Di conseguenza, l'IA riproduce e rafforza modelli culturali euro-centrici o sino-centrici, sottorappresentando voci, culture e istanze provenienti da minoranze e aree in ritardo di sviluppo. Tale dinamica contribuisce a mantenere inalterati squilibri geopolitici e di potere consolidati nei secoli tra gruppi etnici, Paesi ricchi e Paesi poveri. È sufficiente pensare all'utilizzo dell'IA nei sistemi di riconoscimento facciale o di selezione del personale per comprendere quanto questo concetto possa essere rilevante.

Riguardo alla trasparenza, l'Unesco (2023) evidenzia che modelli di IA come Chat GPT sono spesso opachi. In letteratura, questo fenomeno è noto come `*black box problem*': gli algoritmi di IA sono come `scatole nere' che operano con meccanismi difficili da interpretare. La mancanza di trasparenza implica che il processo decisionale dell'IA non sia sempre comprensibile, rendendo complesso per i progettisti sociali individuare errori o discriminazioni. A ciò si aggiunge la problematica dell'affidabilità delle informazioni fornite dall'IA: tali strumenti, non essendo in grado di verificare autonomamente l'accuratezza dei dati generati, possono diffondere informazioni errate o non aggiornate, compromettendo la qualità delle decisioni prese sulla base delle loro risposte.

Infine, è cruciale considerare la protezione della privacy e dei dati personali, un aspetto fondamentale nella progettazione sociale che utilizza l'IA generativa. La progettazione sociale spesso si fonda sulla raccolta, analisi e utilizzo di dati sensibili riguardanti comunità e cittadini, tra cui informazioni sulle condizioni socioeconomiche, la salute, gli aspetti culturali. Senza adeguate misure di sicurezza e un attento controllo, vi è il rischio che questi dati vengano utilizzati in modo improprio, esposti a violazioni o sfruttati per fini discriminatori. Inoltre, la mancanza di trasparenza nei processi decisionali automatizzati può rendere difficile per gli utenti comprendere come vengano trattate le loro informazioni e quali siano i criteri utilizzati per prendere decisioni che li riguardano. Ciò potrebbe creare un contesto in cui le persone non hanno alcun controllo sui propri dati e sono soggette a decisioni opache, potenzialmente lesive dei loro diritti fondamentali. Brynjolfsson e McAfee (2017) sottolineano l'importanza del coinvolgimento umano nel controllo e nella supervisione dell'IA, piuttosto che un affidamento cieco ai sistemi automatici. Un approccio equilibrato, che combini automazione e supervisione umana, è essenziale per garantire decisioni più eque e inclusive, che mettano al centro l'essere umano, la sua dignità e il rispetto dei diritti fondamentali. In questa prospettiva, Floridi (2022) riprende il principio kantiano secondo cui l'essere umano deve essere sempre considerato un fine e mai un mezzo. L'IA deve essere dunque modellata e governata con lucidità ed etica, perché sia al servizio del bene sociale e adattata ai bisogni umani. A tal fine, la responsabilità e l'autonomia delle persone risultano fondamentali per evitare che le persone stesse diventino parte del meccanismo con un ridotto ruolo decisionale.

Alla luce di queste considerazioni, per un uso etico dell'IA nella progettazione sociale, i progettisti sociali devono assumersi la responsabilità nei confronti delle comunità e dei beneficiari, preservando allo stesso tempo la propria autonomia nei processi decisionali. Ciò richiede l'acquisizione di competenze sulla governance dell'IA e l'adozione di strategie per mitigare i rischi etici. Un metodo efficace potrebbe essere l'adozione di pratiche partecipative che coinvolgano diversi stakeholder: rappresentanti delle comunità, organizzazioni non profit, istituzioni ed esperti di etica e IA, insieme potrebbero definire criteri chiari per l'uso dell'IA nei progetti sociali e garantire che essa non amplifichi discriminazioni e disuguaglianze. Inoltre, per ridurre le distorsioni ed evitare sottorappresentazioni, i progettisti dovrebbero garantire che i dati utilizzati per l'addestramento siano diversificati e rappresentativi della popolazione. Sarebbe utile implementare meccanismi di monitoraggio costante dei modelli di IA, per migliorare la trasparenza e l'affidabilità dei risultati attraverso analisi qualitative dei dati, revisioni critiche delle risposte fornite dall'IA, sondaggi o focus group che coinvolgano comunità e beneficiari per discutere e identificare eventuali problemi o distorsioni, e confronti con best practice internazionali. Le azioni per un uso etico dell'IA nella progettazione sociale trovano riferimento negli standard etici e normativi stabiliti a livello internazionale ed europeo. In quanto agli standard etici, le raccomandazioni dell'Unesco (2022) e le linee guida etiche della Commissione europea (2019) promuovono trasparenza, equità

e non discriminazione, ponendo particolare attenzione alla protezione dei gruppi vulnerabili. Riguardo alla privacy, raccomandano la conformità al GDPR (Regolamento UE 2016/679)[7], che implica da parte del progettista la responsabilità di informare le persone coinvolte nei progetti sociali in merito a come e perché l'IA viene usata, garantendo trasparenza e tutela dei dati.

Infine, gli standard etici internazionali affermano il principio della responsabilità e del controllo umano che, applicato alla progettazione sociale, suggerisce che le decisioni debbano rimanere sotto la supervisione del progettista sociale, evitando di delegare scelte eticamente rilevanti esclusivamente all'IA. Sul piano normativo, il GDPR tutela i dati personali e riconosce il diritto a contestare decisioni automatizzate (art. 22; considerando 71). Inoltre, con l'approvazione dell'AI Act (Regolamento UE 2024/1689)[8], il primo quadro giuridico globale sull'IA, sono stati introdotti obblighi stringenti in materia di trasparenza, responsabilità e utilizzo etico. I progettisti devono quindi assicurarsi che le tecnologie impiegate rispettino questi standard, il che comporta un costante aggiornamento normativo e una maggiore complessità gestionale.

Da non dimenticare, infine, il Codice di condotta del progettista sociale: "Il progettista sociale ispira la sua attività professionale al perseguimento dell'utilità sociale dei beneficiari, nel rispetto e in accordo alla mission e vision organizzative dell'ente per cui opera o con cui collabora, in conformità agli scopi di solidarietà sociale iscritti nella Costituzione Italiana e/o alle Convenzioni ONU sui diritti dell'Uomo, sui diritti dell'Infanzia, sui diritti dei Lavoratori migranti, sui diritti delle Persone con Disabilità, sull'eliminazione di ogni forma di discriminazione, della Donna e in conformità a ogni altra Convenzione internazionale già promulgata o ad altre Convenzioni internazionali che riconoscano e proclamino nuovi diritti" (Associazione italiana progettisti sociali 2013, 2).

In conclusione, un approccio critico e umanistico, combinato con il rispetto degli standard etici e normativi, è essenziale per integrare l'IA nella progettazione sociale in modo equo e sostenibile. L'IA deve essere concepita non come un fine, ma come uno strumento al servizio della dignità umana. In questo contesto, è fondamentale che il progettista assuma il ruolo di 'ciberneta' (dal greco *kybernetes*), ovvero di 'timoniere' dell'innovazione tecnologica garantendo un controllo umano consapevole e responsabile per orientare l'uso dell'IA verso il bene comune.

Conclusioni

L'Intelligenza artificiale generativa sta trasformando in modo significativo il processo di progettazione sociale, influenzando ogni fase, dalla creazione dei progetti all'analisi degli impatti generati dagli interventi. Sebbene sia ancora prematuro formulare previsioni dettagliate sull'impatto di questi cambiamenti, la tecnologia è in continua evoluzione e le possibilità future sono incerte. Tuttavia, ciò che sembra chiaro è che stiamo entrando in una fase di maggiore interconnessione e pervasività, accelerata dallo sviluppo degli agenti virtuali. Ci troviamo ora a un bivio: da un lato, accettare passivamente i cambiamenti imposti dalla tecnologia, cercando di sfruttare al massimo i benefici che essa offre; dall'altro, ridisegnare i processi e le pratiche per governare il cambiamento, concentrandosi sull'utilizzo della tecnologia dove può davvero fare la differenza, senza perdere il significato fondamentale del ruolo del progettista sociale. La risposta a questa sfida dipenderà probabilmente dall'equilibrio che riusciranno a trovare da un lato gli enti finanziatori (pubblici e privati) e dall'altro la comunità dei progettisti sociali. Per quanto concerne gli enti finanziatori, un aspetto cruciale riguarda i criteri e gli strumenti di valutazione. Attualmente, la valutazione di un progetto si concentra principalmente sulla qualità e completezza della documentazione scritta, che spesso è redatta da una singola persona, senza coinvolgere pienamente la comunità. In futuro, se le decisioni di approvazione, i riconoscimenti delle spese e la verifica dei risultati non si baseranno più

7 Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (Regolamento generale sulla protezione dei dati) (Testo rilevante ai fini del SEE), in GU L 119 del 4.5.2016.

8 Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i Regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (Regolamento sull'intelligenza artificiale), in GU L 1689 del 12.07.2024.

esclusivamente sui documenti, ma richiederanno un dialogo diretto con i destinatari e gli operatori, l'uso dell'IA potrebbe rivoluzionare il ruolo del progettista sociale, rendendolo più centrato sulla partecipazione attiva e il coinvolgimento delle persone.

Ciò che distingue un progetto sociale da un progetto tecnico tradizionale è che quest'ultimo ha un risultato misurabile e oggettivo, mentre nei progetti sociali le persone e le loro esperienze sono al centro. Non si tratta solo di ottenere un prodotto finale, ma di promuovere cambiamenti nelle relazioni e nelle dinamiche sociali. Se l'interesse per i progetti sociali si concentrasse principalmente sulla qualità dei processi attivati e sulla soddisfazione dei beneficiari, piuttosto che sulla qualità dei documenti prodotti, assisteremmo a una trasformazione profonda della progettazione sociale. In fase di progettazione, non sarebbe sufficiente raccogliere dati sul contesto e sui problemi, ma sarebbe altrettanto fondamentale creare partecipazione, motivazione e comprensione tra gli operatori e i destinatari. Se la valutazione del progetto comprendesse questi aspetti, il progettista sociale non sarebbe più solo un redattore di documenti, ma un facilitatore e manager dei processi collaborativi. L'IA potrebbe offrire un importante valore aggiunto nell'analisi dei dati, nell'elaborazione dei contenuti e nella generazione di idee creative. Tuttavia, il progettista sociale rimarrebbe il punto di riferimento per la gestione delle dinamiche relazionali, per il coinvolgimento dei soggetti interessati e per l'attivazione di processi di trasformazione sociale. Per la comunità dei progettisti sociali, l'approccio nei confronti degli strumenti di IA diventa, quindi, determinante. Se i progettisti riusciranno a spostare il focus dalla mera produzione di documenti verso il porsi le domande giuste e l'estrarre valore dalle nuove tecnologie, il governo del processo rimarrà nelle mani degli esseri umani. Sarà necessario affinare la capacità critica di fronte a elaborati sempre più sofisticati e articolati, che potrebbero nascondere *bias* o non tenere conto delle peculiarità dei contesti locali. Il progettista sociale dovrà interagire costantemente con l'IA, precisando gli input e migliorando gli output in un processo di ottimizzazione continua, con una visione strategica ben ancorata ai bisogni sociali.

Il rischio che non si realizzi questa evoluzione è alto. Già oggi assistiamo a progetti che spesso risultano slegati dal contesto e dalle persone che vi partecipano. Se l'IA rende più facile produrre progetti complessi, il rischio è che questi siano esteticamente e tecnicamente impeccabili, ma distanti dalle reali necessità delle comunità. La sfida per i progettisti sociali sarà mantenere il focus sull'essenza del loro lavoro: le persone e le comunità, evitando la tentazione di semplificare il processo di progettazione in nome dell'efficienza. Infine, va sottolineato che, dato il rapido sviluppo degli strumenti di IA, è difficile prevedere con certezza come evolverà il panorama della progettazione sociale. Se scrivessimo questo articolo tra un anno, probabilmente avremmo aggiornato le nostre riflessioni in modo significativo. È quindi fondamentale continuare a interrogarci, adattarci e cercare nuovi equilibri, sempre fedeli alla nostra missione sociale e al ruolo fondamentale del progettista sociale.

Bibliografia

Abbass H.A. (2019), Social Integration of Artificial Intelligence: Functions, Automation Allocation Logic and Human-Autonomy Trust, *Cognitive Computation*, 11, pp.159-171

Alshaikhi A., Khayyat M. (2021), An investigation into the Impact of Artificial Intelligence on the Future of Project Management *2021 International Conference of Women in Data Science at Taif University (WiDSTaif), Taif, Saudi Arabia,* pp.1-4 <DOI 10.1109/WiDSTaif52235.2021.9430234>

Arnstein S.R. (1969), A Ladder of Citizen Participation*, Journal of the American Planning Association*, 35, n.4, pp.216-224

Associazione italiana Progettisti sociali (2013), *Codice di condotta*, Roma, Associazione italiana progettisti sociali

Banavar G. (2016), *Learning to trust artificial intelligence systems*, IBM Report, Somers, NY, IBM

Bankhwal M., Bisht A., Chui M., Roberts R., van Heteren A. (2024), *AI for social good: Improving lives and protecting the planet,* McKinsey e Company Report*,* Chicago*,* McKinsey e Company

Barcaui A., Monat A. (2023), Who is better in project planning? Generative artificial intelligence or project managers? *Project Leadership and Society*, 4, n.1, 100101

Benasayag M., Pennisi A. (2024), *ChatGPT non pensa (e il cervello neppure)*, Milano, Jaca Book

Benedetto XVI (2016), *Discorso di Sua Santità Benedetto XVI ai partecipanti al Convegno promosso dal Partito Popolare Europeo*, Città del Vaticano, Libreria Editrice Vaticana

Berger P., Luckmann T. (2016), *The social construction of reality*, in Longhofer W., Winchester D. (eds.), *Social theory re-wired*, New York, Routledge, pp.110-122

Booyse D., Scheepers C.B. (2023), Barriers to adopting automated organisational decision-making through the use of artificial intelligence, *Management Research Review*, 47, n.1, pp.64-85

Brynjolfsson E., Mcafee A. (2017), The Business of Artificial Intelligence, in Brynjolfsson E., Mcafee A. (eds.), *Artificial Intelligence, For Real*, Brighton, MA, Harvard Business Publishing Corporation, pp.3-11

Cascino G. (2024), Metodi e tecniche di progettazione sociale e territoriale, *Corso di Laurea Magistrale in Scienze Sociali per lo Sviluppo Sostenibile*, *A.A. 2024/2025*, Dipartimento di Scienze dell'Uomo e della Società, Enna, Università degli Studi di Enna "Kore"

Chiodi M., Costa G., Cucchetti S., Fontana P., Picozzi M., Reichlin M. (2012), Valori non negoziabili: una categoria che fa discutere, *Aggiornamenti Sociali*, 9-10, pp.656-657

Grice H.P. (1989), *Studies in the Way of Words*, Cambridge, MA, Harvard University Press

Commissione europea (2019), *Orientamenti etici per un'IA affidabile*, Bruxelles, Commissione europea <doi:10.2759/640340>

Di Troia S., Parli V., Pava J.N., Badi-Uz-Zaman H., Fitzsimmo K. (2024), *Inspiring Action: Identifying the Social Sector AI Opportunity Gap*, Working Paper, Stanford, The Stanford Institute for Human-Centered Artificial Intelligence (HAI)

Eco U. (1999), *I Limiti dell'Interpretazione*, Milano, Bompiani

European Union Agency for Fundamental Rights (FRA) (2022), *Bias in Algorithms. Artificial Intelligence and Discrimination, FRA Report 2022*, Luxembourg, Publications Office of the European Union

Floridi L. (2022), *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, Scienze e idee n.340, Milano, Raffaele Cortina Editore

Formez (2002), *Project Cycle Management. Manuale per la Formazione*, Roma, Presidenza del Consiglio dei Ministri, Dipartimento della Funzione pubblica

Gînguță A., Ștefea P., Noja G.G., Munteanu V.P. (2023), Ethical Impacts, Risks and Challenges of Artificial Intelligence Technologies iBusiness Consulting: A New Modelling Approach Based on Structural Equations, *Electronics*, 12, n.6, 1462

Glikson E., Woolley A.W. (2020), Human Trust in Artificial Intelligence: Review of Empirical Research, *Academy of Management Annals*, 14, pp.627-660

Guido R. (2024), Intelligenza Artificiale e psicologia: l'impatto sulle nostre relazioni sociali e nei contesti organizzativi, *Bnews Università Milano-Bicocca,* 17 luglio

Husserl E. (2008), *La crisi delle scienze europee e la fenomenologia trascendentale*, Tascabili n.24, Milano, Il Saggiatore (Opera originale pubblicata nel 1954)

Innes J.E., Booher D.E. (2010), *Planning with Complexity: An Introduction to Collaborative Rationality for Public Policy*, London, Routledge

Kahneman D., Tversky A. (1979), Prospect Theory: An analysis of decision under risk, *The Econometric Society*, 47, n.2, pp.263-291

Kamila M.K., Jasrotia S.S. (2025), Ethical issues in the development of artificial intelligence: recognizing the risks, *International Journal of Ethics and Systems*, 41, n.1, pp.45-63

Ledwith M. (2016), *Community Development in Action. Putting Freire into Practice*, Bristol, Policy Press

Leone L., Prezza M. (2003), *Costruire e valutare i progetti nel sociale*, *Manuale operativo per chi lavora su progetti in campo sanitario, sociale, educativo e culturale*, L'operatore sociale nella professione n.5, Milano, Franco Angeli

Manzini E. (2015), *Design, When Everybody Designs. An Introduction to Design for Social Innovation*, Cambridge (MA), MIT Press

Mazzucato A., Bortolami A., Caravello G. (2025), Intelligenza Artificiale Generativa e Progettazione sociale: opportunità e sfide, *Welforum.it,* 24 gennaio

Mert K. (2024), Informed empathy in design practice: Exploring designers'empathy in reference to unfamiliar user groups*, Ph.D. - Doctoral Program,* Ankara, Middle East Technical University

Mohamed S., Png M.-T., Isaac W. (2020), Decolonial AI. Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence, *Philosophy* e *Technology,* 33, pp.659-684

Norman D.A. (2019), *La caffettiera del masochista. Il design degli oggetti quotidiani*, Firenze, Giunti

Paolo G., Gonzales-Billandon J., Kégl B. (2024), A call for embodied AI, *arXiv.org* <https://doi.org/10.48550/arXiv.2402.03824>

Plebe A. (2019), Deep learning vs scienze cognitive: così l'intelligenza si è disconnessa dal corpo, *agendadigitale.eu*, 22 ottobre

Qian Y., Siau K.L., Nah F.F. (2024), Societal Impacts of Artificial Intelligence: Ethical, Legal, and Governance Issues, *Societal Impacts*, 3, n.3, 100040

Ryan M., Stahl B.C. (2020), Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications, *Journal of Information, Communication and Ethics in Society*, 19, n.1, pp.61-86

Sicora A., Pignatti A. (2024), *Progettare sociale. Progettazione e finanziamenti europei per i servizi sociali ed educativi*, Santarcangelo di Romagna, Maggioli Editore

Simon H.A. (1985), *Causalità, razionalità, organizzazione*, Bologna, il Mulino

Simon H.A. (1947), *Administrative behavior. A Study of Decision-Making Processes in Administrative Organization*, New York, Macmillan

Siza R. (2018), *Manuale di progettazione sociale*, Sociologia, cambiamento e politica sociale n.10, Milano, Franco Angeli

Smith E.R., Gün R.S. (2004), Socially situated cognition: Cognition in its social context, *Advances in experimental social psychology*, 36, pp.57-121

Unesco (2023), *Foundation models such as ChatGPT through the prism of the UNESCO Recommendation on the Ethics of Artificial Intelligence*, Paris, Unesco

Unesco (2022), *Recommendation on the Ethics of Artificial Intelligence*, Paris, Unesco

Wang J., Zhang C., Li J., Ma Y., Niu L., Han J., Peng Y., Zhu Y., Fan L. (2024), Evaluating and Modeling Social Intelligence: A Comparative Study of Human and AI Capabilities, *arXiv.org* <https://doi.org/10.48550/arXiv.2405.11841>

Weber M. (2001), *L'etica protestante e lo spirito del capitalismo*, Milano, Rizzoli (Opera originale pubblicata nel 1905)

Zajko M. (2022), Artificial intelligence, algorithms, and social inequality: Sociological contributions to contemporary debates, *Sociology Compass*, 16, n.3, e12962

Zhuo T.Y., Huang Y., Chen C., Xing Z. (2023), Exploring AI Ethics of Chatgpt: A Diagnostic Analysis, *arXiv.org* <DOI:10.48550/arXiv.2301.12867>

Alberto Bortolami
albeborto@gmail.com
Direttore generale della Cooperativa sociale NOIGroup e membro del Consiglio direttivo dell'Associazione italiana progettisti sociali (APIS). È laureato in Economia e Direzione aziendale ed ha conseguito due master, uno in Gestione etica d'azienda e l'altro in Counselling educativo. Ha lavorato in Fondazione Cariparo, in cui ha gestito i principali progetti nel sociale. Successivamente ha ricoperto il ruolo di Direttore in Fondazione Oggi e Domani ETS, spin-off di Fondazione Cariparo sulla tematica del 'Dopo di noi'. Ha sviluppato percorsi formativi in collaborazione con vari enti, tra cui l'Università degli Studi di Padova.

Francesco Capone
fcapone68@gmail.com
Laureato in Economia con una tesi sulla gestione dei volontari nelle organizzazioni non profit, si è successivamente formato alla scuola di Carlo Borgomeo e Giuseppe De Rita. Membro del Consiglio direttivo dell'Associazione italiana progettisti sociali (APIS), si occupa da oltre 20 anni di sviluppo locale, creazione e sviluppo di impresa, progettazione sociale e valorizzazione a fini sociali di beni e aziende confiscate alle mafie. Fra le pubblicazioni si segnalano: *La Politica Educata. Una proposta alla formazione alla politica della Diocesi di Lecce*, Rubettino, 2021; *Il terzo settore nel Sud d'Italia*, Edizioni Gruppo Abele, 2000.

Giusy Caravello
giusycaravello@hotmail.com
Dal 2005 si occupa di sordità e sordocecità infantili, progettando interventi inclusivi presso Cabss Onlus. Membro del Consiglio direttivo dell'Associazione italiana progettisti sociali (APIS), è laureata in Scienze politiche presso l'Università di Messina e specializzata in Management del Terzo settore (Università S. Tommaso d'Aquino di Roma). Fra le pubblicazioni recenti: (con Fadda S., Cidronelli M., Harripersad L.), La Terapia cognitivo-comportamentale accessibile ai bambini sordociechi, *Quaderni Di Psicoterapia Cognitiva*, 2023.

Antonio Finazzi Agrò
a.finazziagro@gmail.com
Vicepresidente e formatore presso l'Associazione italiana progettisti sociali (APIS). Ha collaborato in qualità di esperto alla redazione della Norma tecnica UNI 11746:2019 (progettista sociale). Ha operato come consulente del Terzo settore e PA, con specializzazione in progettazione e project management di interventi sociali e cooperazione allo sviluppo. Da giugno 2020 è Presidente de La Nuova Arca Società cooperativa sociale. Fra le pubblicazioni più recenti si segnalano: La crisi del lavoro sociale, tra fine della narrazione sussidiaria e working poor, *Osservatorio nazionale sulle politiche sociali, Welforum.it*, 2024; Il Principio Generativo tra eccedenza, interiorità e speranza, *Quaderni Genialis*, n.3, 2025.

Federico Maggiora
f.maggiora@accademiamm.it
Progettista esperto in innovazione sociale, attualmente si occupa di ESG all'interno della Direzione Pianificazione e controllo di gestione della Divisione Banca dei Territori di Intesa Sanpaolo. È stato Coordinatore di progetti della Fondazione per l'Innovazione del Terzo settore Banca Prossima e membro dello staff del CEO dello stesso gruppo bancario. È Presidente della Fondazione Accademia Maurizio Maggiora ETS ed esperto in business modelling sociale e in accompagnamento allo sviluppo professionale di progettisti sociali. È membro del Consiglio direttivo e Presidente del Comitato scientifico dell'Associazione italiana progettisti sociali (APIS).

Annaleda Mazzucato
annaleda.mazzucato@mondodigitale.org
Project manager ed European fundings coordinator presso Fondazione Mondo Digitale, Ambasciatore Erasmus+ EDA e valutatore Indire, PhD presso l'Università degli Studi di Napoli Federico II nei settori scientifico disciplinari Psicologia e Scienze dell'educazione e membro del Consiglio direttivo dell'Associazione italiana progettisti sociali (APIS), si occupa del ruolo delle tecnologie nei processi di innovazione sociale e dell'apprendimento. Ha conseguito un Master in progettazione europea presso l'Università degli Studi di Padova, il Certificate in Fundraising Management presso il London Institute of Fundraising e la Certificazione Project Management Professional presso il Project Management Institute.

Intelligenza vs opacità istituzionale
La comunicazione pubblica alla sfida dell'Intelligenza artificiale generativa nell'Unione europea

Sara Pane
Sapienza Università di Roma
Università di Torino

L'IA generativa trasforma la comunicazione pubblica europea, offrendo servizi più personalizzati, ma sollevando sfide etiche e competenziali. Questo studio analizza criticamente opportunità e rischi, proponendo il paradigma della 'comunicazione pubblica 4.0'. Attraverso i casi GPT@EC ed Eurodesk Chatbot, si esplora il potenziale dei chatbot istituzionali, evidenziando la necessità di un approccio umano-centrico basato su trasparenza, inclusività e responsabilità.

Generative Artificial intelligence is reshaping public communication in Europe by enhancing service personalization while introducing significant ethical, technical, and governance challenges. This study provides a critical assessment of opportunities and risks, proposing the paradigm of 'Public Communication 4.0'. Through the analysis of GPT@EC and the Eurodesk Chatbot, it investigates the potential of institutional chatbots, emphasizing the need for a human-centered approach based on transparency, inclusiveness, and ethical accountability.

DOI: 10.53223/Sinappsi_2025-02-5

Citazione
Pane S. (2025), Intelligenza vs. opacità istituzionale. La comunicazione pubblica alla sfida dell'Intelligenza artificiale generativa nell'Unione europea, *Sinappsi*, XV, n.2, pp.62-74

Parole chiave
Comunicazione pubblica
Intelligenza artificiale
Unione europea

Keywords
Public communication
Artificial intelligence
European Union

Introduzione

L'Intelligenza artificiale (d'ora in poi indicata con l'acronimo IA) si presenta come un catalizzatore di profonde trasformazioni per il settore pubblico (Laux *et al*. 2024). Se gestita in modo appropriato, questa tecnologia ha il potenziale per rivoluzionare sia l'erogazione dei servizi pubblici sia le interazioni tra cittadini e istituzioni (Androutsopoulou *et al*. 2019). Dalla semplificazione delle procedure burocratiche all'ottimizzazione delle risorse, l'adozione di strumenti basati sull'IA rappresenta un'opportunità strategica per migliorare l'efficienza dei processi e innovare le strutture organizzative delle pubbliche amministrazioni (Ng *et al*. 2021). Tuttavia, accanto a queste opportunità emergono anche delle sfide, soprattutto sul piano etico (Floridi 2021). Le ripercussioni sulla privacy degli utenti e sull'occupazione nel pubblico impiego, insieme all'imprevedibilità delle logiche algoritmiche e alla trasparenza dei dati, rappresentano alcune delle principali criticità legate all'adozione dell'IA nel settore pubblico (Floridi 2021; Susar e Aquaro 2019). In particolare, nell'ambito specifico della comunicazione pubblica che, in generale, si riferisce

alla sfera degli scambi tra cittadini e istituzioni (Lovari e Ducci 2022). Inizialmente concepita secondo una logica verticale e unidirezionale, la comunicazione pubblica ha subito una profonda trasformazione nel corso del tempo, favorita dalla transizione digitale, che ha consentito una maggiore apertura verso il cittadino sempre più considerato come un utente (Lovari *et al*. 2020). La rapida diffusione dell'IA, in particolare quella generativa, preannuncia un ulteriore cambiamento di paradigma e apre la strada a un modello di 'comunicazione pubblica 4.0' in cui il cittadino-utente diventa coprotagonista insieme agli algoritmi e alle piattaforme (OECD 2021). Nell'Unione europea (di qui in poi UE), il Regolamento sull'IA (Parlamento europeo e Consiglio europeo 2024) ha introdotto un quadro normativo senza precedenti, volto a promuovere uno sviluppo dell'IA che ponga al centro il benessere e i diritti dei cittadini, in linea con un approccio umano-centrico. Questa regolamentazione, unita alla progressiva integrazione degli strumenti di IA nelle pratiche di comunicazione pubblica, e più in generale nel settore pubblico, fa del contesto europeo un punto di riferimento privilegiato per osservare e guidare la transizione verso un uso responsabile e sostenibile dell'IA (Floridi 2021; Laux *et al*. 2024; Mökander *et al*. 2022). Ciononostante, gli studi che affrontano la questione da una prospettiva duplice, che integra la teoria con l'esperienza più pratica del policy making, sono ancora pochi. Inserendosi in questo gap di ricerca, il presente contributo esplora l'uso dell'IA generativa nel settore pubblico europeo, concentrandosi in particolare nell'ambito della comunicazione pubblica istituzionale. Attraverso un'analisi critica della letteratura e il riferimento ai dati della recente indagine condotta dall'Osservatorio europeo Public Sector Tech Watch sull'impiego dell'IA generativa nel settore pubblico europeo, questo studio mira a offrire una comprensione approfondita delle opportunità e delle sfide legate all'adozione di questa tecnologia, con particolare attenzione alle implicazioni per le attività di comunicazione pubblica a livello istituzionale (Grimmelikhuijsen e Tangi 2024). Le domande di ricerca che hanno guidato l'analisi sono le seguenti: *quali sono le principali opportunità e sfide derivanti dall'uso dell'IA generativa nel settore pubblico nell'UE? Nel campo specifico della comunicazione pubblica istituzionale europea, l'IA generativa (in particolare attraverso i 'chatbot istituzionali') può migliorare l'interazione tra istituzioni e cittadini?* Nel tentativo di fornire una risposta puntuale a questi interrogativi, particolare rilievo è attribuito ai 'chatbot istituzionali', considerati strumenti strategici per favorire l'innovazione e migliorare la qualità dell'interazione tra le istituzioni pubbliche e i cittadini nell'Unione, soprattutto sul piano della comunicazione. In questo contesto, vengono analizzati due casi studio: uno relativo alla comunicazione interna e l'altro alla comunicazione esterna dell'UE (Androutsopoulou *et al*. 2019; Makasi *et al*. 2020; van Noordt e Misuraca 2022).

1. Trasformazioni e sfide della comunicazione pubblica: dal modello unidirezionale alla partecipazione attiva dei cittadini

Storicamente, la comunicazione pubblica è stata concepita come una funzione puramente trasmissiva all'interno delle organizzazioni, caratterizzata da un'impostazione verticale, gerarchica e unidirezionale, in cui le amministrazioni interagivano con i cittadini, spesso considerati come destinatari passivi (Lovari *et al*. 2020). Zémor (1995) definisce la comunicazione pubblica come una forma di comunicazione formale finalizzata allo scambio e alla condivisione di informazioni di interesse pubblico, al mantenimento dei legami sociali, con la responsabilità affidata a istituzioni pubbliche o organizzazioni che perseguono finalità di interesse collettivo. Per lungo tempo, l'assenza di un'adeguata istituzionalizzazione degli uffici di comunicazione all'interno degli organigrammi del settore pubblico ha determinato che amministrazioni e istituzioni fossero prevalentemente comunicate attraverso le narrazioni promosse dai politici e dai media (Lovari *et al*. 2020). Di conseguenza, la comunicazione pubblica è stata spesso erroneamente confusa con i discorsi degli attori politici, i quali la utilizzavano principalmente per promuovere le proprie agende (OECD 2021). Questo ha portato a una percezione distorta, secondo cui la comunicazione pubblica fosse prevalentemente finalizzata a scopi propagandistici e persuasivi nei confronti dell'opinione pubblica. I media hanno contribuito a questa rappresentazione, dando ampia copertura alle questioni pubbliche prediligendo criteri di notiziabilità focalizzati su scandali, disservizi e cattiva gestione delle risorse condivise (Lovari *et al*. 2020). Ciò ha supportato la diffusione generalizzata di un frame di inefficienza, influenzando negativamente

l'opinione collettiva e il modo in cui la popolazione concepisce il settore pubblico. Questo modello di comunicazione ha cominciato a evolversi a causa di vari fattori socio-tecnici e delle riforme progressive del settore pubblico, che hanno interessato le società occidentali a partire dalla fine degli anni Settanta del Novecento (Lovari *et al*. 2020). Tali trasformazioni hanno comportato una profonda riorganizzazione delle strutture e dei processi operativi delle pubbliche amministrazioni, favorendo la loro graduale apertura verso l'esterno. Sul piano organizzativo, il passaggio dal cosiddetto modello 'burocratico' delle pubbliche amministrazioni, tradizionalmente descritto dal sociologo tedesco Max Weber in termini di rigidità, verticalità e impersonalità, al modello del *New Public Management*, si è inizialmente verificato nei Paesi anglosassoni, seguendo le innovazioni introdotte da Ronald Reagan negli Stati Uniti e Margaret Thatcher nel Regno Unito. Solo successivamente, a partire dagli anni Novanta, si è diffuso nei Paesi dell'Europa continentale per poi estendersi al resto dell'Unione (Pollitt *et al*. 2007). Ispirato ai principi del settore privato e dell'economia di mercato, questo nuovo paradigma ha orientato le istituzioni verso una maggiore efficienza operativa, una più ampia trasparenza amministrativa e un'attenzione crescente nei confronti del cittadino (Pollitt e Dan 2011). Tale evoluzione ha avuto effetti significativi anche nel contesto della comunicazione pubblica, promuovendo l'ascolto attivo e il meccanismo del feedback (Lovari e Ducci 2022). A sua volta, ciò ha favorito la transizione verso un modello di comunicazione pubblica meno rigido e più bidirezionale, caratterizzato da una maggiore apertura delle amministrazioni nei confronti dei cittadini. Questo processo è stato ulteriormente accelerato dalla rivoluzione digitale, avviata con la diffusione di Internet, le cui origini risalgono alla rete di computer Arpanet, sviluppata negli Stati Uniti nel 1969. Per ragioni epistemologiche e di pertinenza con l'oggetto di studio, il presente contributo non si concentra sulla ricostruzione storica della transizione digitale, bensì sulle sue implicazioni e ricadute più rilevanti per le organizzazioni del settore pubblico. La trasformazione digitale ha, infatti, fornito le risorse cognitive e materiali per modernizzare la comunicazione pubblica ed espandere la cultura orientata al cittadino nel settore pubblico. Istituzioni e pubbliche amministrazioni hanno normalizzato la sperimentazione di nuove interfacce digitali e l'istituzionalizzazione di blog e social media come canali di comunicazione (Lovari e Ducci 2022). All'interno di questo rinnovato contesto, la quantità di comunicazione pubblica è aumentata. Tuttavia, tale incremento quantitativo non è stato accompagnato da un corrispondente miglioramento della qualità. Per lo meno, non in termini assoluti. Gli ambienti digitali hanno, infatti, favorito la proliferazione di informazioni fuorvianti e non sempre verificate, contribuendo alla diffusione della disinformazione e influenzando negativamente la percezione delle questioni pubbliche in un contesto già permeato da fragilità (Smillie e Scharfbillig 2024). Si considerino, ad esempio, le campagne di comunicazione pubblica implementate tra il 2020 e il 2021 a sostegno delle misure istituzionali adottate per contrastare l'infodemia legata alla pandemia da Covid-19 (OECD 2021). In questo contesto, le istituzioni – che, come si vedrà in seguito, sono tra i principali attori della comunicazione pubblica – sono state chiamate a progettare e implementare strategie di comunicazione che, oltre ad informare i cittadini sull'evoluzione della pandemia, contrastassero anche i flussi di disinformazione e misinformazione presenti nello spazio pubblico con ripercussioni negative anche sul piano della salute collettiva. La gestione della crisi è stata fortemente influenzata dalla capacità delle autorità pubbliche di incoraggiare i cittadini ad adottare comportamenti specifici a beneficio dell'intera collettività (OECD 2021). Per questo motivo, gli esperti concordano nel ritenere che l'emergenza sanitaria da Covid-19 abbia rappresentato un vero e proprio punto di svolta nell'ambito della comunicazione pubblica (Lovari e Belluati 2023; OECD 2021). A partire da questa ricostruzione, seppure semplificata, emerge come la comunicazione pubblica abbia attraversato un processo di trasformazione significativo, che ha portato a una ridefinizione del suo ambito e del suo ruolo all'interno delle società contemporanee. Attualmente, la comunicazione pubblica è prodotta prevalentemente da attori istituzionali: ministeri, agenzie governative, pubbliche amministrazioni e altri soggetti coinvolti nella trasmissione di contenuti di rilevanza pubblica. La comunicazione pubblica è definita come un insieme di relazioni professionali finalizzate a "elaborare e diffondere contenuti informativi che esprimano l'identità di un'istituzione" (Nieto 2006, 58). Essa abbraccia un ampio spettro di attività, che includono sia la comunicazione interna che quella esterna. La comunicazione interna riguarda il flusso di

informazioni tra i vari dipartimenti di un'organizzazione, facilitando la trasmissione di dati. La comunicazione esterna, invece, si articola su due fronti: da un lato, si rivolge ai cittadini con l'obiettivo di fornire informazioni e sensibilizzare su tematiche di interesse collettivo. In questo senso, facilita l'ascolto e la risposta istituzionale alle esigenze dei cittadini e promuove una governance più inclusiva ed efficace. Dall'altro lato, la comunicazione pubblica si rivolge ai media, in particolare a quelli di massa, con la finalità di comunicare le attività e i risultati conseguiti, nonché di contribuire alla costruzione e alla promozione dell'immagine istituzionale all'esterno. In questa prospettiva, la comunicazione pubblica è strategica (OECD 2021). Nonostante il suo potenziale intrinseco, la comunicazione pubblica continua a svolgere una funzione ausiliaria nel processo di policy making e il limitato coinvolgimento dei comunicatori durante la progettazione delle politiche pubbliche spesso porta alla trasmissione di contenuti preimpostati e poco efficaci (Canel e Luoma-aho 2018; Canel e Sanders 2013; Lovari *et al*. 2020; Macnamara 2017). Un'interessante prospettiva di analisi emerge nel contesto europeo, dove i recenti progressi tecnologici, principalmente legati alla rapida diffusione dell'IA, sembrano aver creato le condizioni per un cambiamento paradigmatico nella comunicazione pubblica. Questo processo segna l'ingresso in una nuova fase, caratterizzata dall'ulteriore trasformazione dei modelli di interazione tra cittadini e istituzioni, dando vita a un nuovo paradigma di 'comunicazione pubblica 4.0' (Lovari e Belluati 2023). Nel contesto dell'Unione, la trasparenza amministrativa si configura oggi come una condizione strutturale per rafforzare la fiducia tra cittadini e istituzioni, ma anche come garanzia essenziale per l'affidabilità dell'IA nelle pubbliche amministrazioni. Non si tratta più soltanto di rendere accessibili documenti o dati, ma di assicurare che le decisioni algoritmiche siano tracciabili, comprensibili e contestabili. Questa prospettiva è pienamente recepita nel Regolamento UE 2024/1689 sull'Intelligenza artificiale (*Artificial intelligence Act*), che stabilisce un quadro giuridico armonizzato per lo sviluppo e l'uso responsabile dell'IA nell'Unione. In particolare, il documento precisa che i sistemi di IA devono essere progettati secondo principi di trasparenza, spiegabilità e supervisione umana, in coerenza con le *Ethics Guidelines for Trustworthy AI* elaborate dalla Commissione. L'articolo 13 del Regolamento introduce obblighi stringenti per i sistemi ad alto rischio, imponendo requisiti di documentazione tecnica, registrazione degli eventi e tracciabilità delle decisioni automatizzate. Tali disposizioni trasformano la trasparenza da principio etico astratto a vincolo operativo, specie nei contesti ad alta sensibilità come la comunicazione pubblica. L'integrazione di sistemi basati su IA, come i chatbot, comporta l'obbligo di informare chiaramente gli utenti sul fatto che stanno interagendo con un'Intelligenza artificiale, sulle modalità di funzionamento e sugli strumenti di tutela disponibili.

Questa nuova forma di trasparenza algoritmica si affianca alla trasparenza amministrativa tradizionale, contribuendo alla costruzione di una 'comunicazione pubblica 4.0': digitale, interattiva, intelligente e responsabile. L'IA non sostituisce i principi della comunicazione pubblica, ma li amplifica e li istituzionalizza a livello sovranazionale, rafforzando la legittimità democratica delle pratiche comunicative dell'Unione europea.

2. La comunicazione pubblica dell'Unione europea: definizione di un concetto multilivello

Con l'avanzare del processo di integrazione europea, avviato dopo la Seconda Guerra mondiale come strumento per prevenire il ripetersi degli eventi drammatici che hanno caratterizzato la prima metà del XX secolo, sono state introdotte delle istituzioni sui generis che hanno determinato la riconfigurazione della tradizionale sovranità politica in chiave sovranazionale. All'interno di questo nuovo sistema, gli Stati membri hanno iniziato a trasferire parte delle loro prerogative agli organismi comuni[1], con conseguenti effetti sugli ordinamenti giuridici nazionali. Di fronte a questa evoluzione, è diventato

1 Gli Stati membri conferiscono competenze all'Unione europea solo quando ricorrono le condizioni giuridiche e funzionali previste dai Trattati, valutando l'efficacia, la necessità e l'appropriatezza dell'intervento comune. Tale conferimento è limitato, reversibile e soggetto a controllo politico e giurisdizionale. Esso si fonda sui principi di attribuzione (l'UE può agire solo nei settori di competenza espressamente conferiti), sussidiarietà (l'intervento dell'Unione è giustificato solo se gli obiettivi non possono essere sufficientemente raggiunti a livello nazionale) e proporzionalità (l'azione europea deve limitarsi a quanto strettamente necessario per raggiungere gli scopi previsti dai Trattati), a garanzia dell'equilibrio tra integrazione e sovranità. Cfr. Manzella (2000).

essenziale informare i cittadini sui cambiamenti socio-politici in atto, con l'obiettivo di accrescerne la consapevolezza riguardo agli sviluppi in corso. Si è aperta, così, la riflessione sulla comunicazione pubblica dell'UE. Le modalità con cui è stata affrontata la questione sono strettamente connesse alle fasi che hanno caratterizzato la costruzione europea. La comunicazione pubblica, inizialmente trascurata e concepita in una logica prevalentemente verticale e unidirezionale, è stata gradualmente riconosciuta come uno strumento utile per sostenere il processo di integrazione europea. A partire dalle prime elezioni a suffragio universale diretto del Parlamento europeo (1979) e, successivamente, con le modifiche all'assetto istituzionale comunitario definite dal Trattato di Maastricht (1992), la comunicazione pubblica europea ha acquisito maggiore autorevolezza ed è stata concepita come una funzione istituzionale riconosciuta (D'Ambrosi 2019). Nell'era post-Maastricht è emersa con maggiore evidenza la necessità di costruire e veicolare messaggi mirati sulle politiche europee, con l'obiettivo duplice di informare i cittadini e stimolare la loro partecipazione attiva nello spazio pubblico che si stava delineando a livello europeo. Qualche anno dopo, durante la prima Presidenza della Commissione europea di José Manuel Durão Barroso (2004-2009), questo trend è stato consolidato. A pochi mesi dal suo insediamento, la Commissione ha rilasciato il Piano d'azione per migliorare la comunicazione in Europa, identificando la comunicazione come uno degli obiettivi strategici da perseguire nel corso della Legislatura (*ibidem*). Riconoscendo i limiti intrinseci dell'approccio prevalente fino a quel momento, caratterizzato da un'attenzione maggiore alle esigenze dell'emittente rispetto a quelle del destinatario, il nuovo modello di comunicazione pubblica europea ha proposto un cambiamento radicale, mettendo il cittadino al centro del processo comunicativo e valorizzando i principi della governance multilivello (Lovari e Belluati 2023). Lo stesso concetto di 'governance multilivello' implica che la comunicazione tra le istituzioni europee e i cittadini non rappresenti più un semplice trasferimento di informazioni dall'alto verso il basso, bensì un processo dialogico e rafforzato nella sua dimensione orizzontale (Ducci 2013). A partire dalla prima decade degli anni Duemila, l'impulso verso il digitale e l'uso intensivo dei social media da parte delle istituzioni europee hanno creato nuove opportunità, sebbene non prive di criticità, per rafforzare ulteriormente la comunicazione pubblica orientata al cittadino-utente (D'Ambrosi 2019). La discesa in campo delle istituzioni europee sulle principali piattaforme di social media ha permesso loro di entrare in diretto contatto con i cittadini, abbattendo le barriere tradizionali dell'accessibilità e rendendo la comunicazione più trasparente e user-friendly. In questo contesto, le istituzioni europee hanno dovuto semplificare il tradizionale linguaggio burocratico, per adattarsi alla logica delle piattaforme digitali e rendere i contenuti più comprensibili e coinvolgenti, ricorrendo a strategie innovative come nel caso dello storytelling digitale. Il susseguirsi di questi sviluppi, qui ricostruito in modo sintetico ma meritevole di un approfondimento più dettagliato, ha portato alla definizione dell'ecosistema comunicativo europeo e alla sua progressiva ridefinizione nel corso del tempo, specie sul fronte della comunicazione esterna. Attualmente, la comunicazione esterna delle istituzioni europee, si configura come un sistema complesso, diversificato e integrato di flussi comunicativi provenienti dalle istituzioni dell'UE e dai suoi organismi (istituzioni, agenzie, centri di ricerca ecc.), dai media degli Stati membri e dai cosiddetti 'media europei' (*Euronews*, *Euractiv*, *Politico Europe*, *Europe by Satellite*). A questi si aggiungono attori come la società civile, i partiti politici europei e le piattaforme digitali (Lovari e Belluati 2023). La funzione della comunicazione pubblica dell'UE rivolta verso l'esterno è duplice. Da un lato, informa i cittadini spiegando e promuovendo i programmi e le politiche dell'Unione. Dall'altro, ne stimola la partecipazione attiva (Pane 2025). Nello scenario contemporaneo, caratterizzato da dinamicità, mutevolezza e incertezza, le istituzioni europee hanno cercato soluzioni innovative per preservare la propria legittimità istituzionale, migliorare l'interazione con i cittadini e offrire prestazioni più efficaci. In questo contesto, l'IA generativa è emersa sia come uno strumento promettente sia come una sfida stimolante che, se adeguatamente sfruttata, può contribuire a raggiungere questi obiettivi (van Noordt e Misuraca 2022).

3. Intelligenza istituzionale: opportunità e sfide dell'Intelligenza artificiale generativa nel settore pubblico nell'Unione europea

Attualmente, il dibattito sull'integrazione dell'IA generativa nel settore pubblico nell'UE è vivo e coinvolge diverse aree di competenza,

che spaziano dall'informatica all'ingegneria, fino alle scienze sociali. Tuttavia, non si tratta di una questione del tutto nuova. Le prime ricerche sull'IA risalgono agli studi congiunti di Warren McCulloch e Walter Pitts (1943) e ai contributi di Alan Turing (1950), avviati negli Stati Uniti a partire dagli anni Quaranta del secolo scorso (Marongiu 2020). In un saggio del 1950, Turing ha formulato per la prima volta il concetto di '*computer*', inteso come una '*thinking machine*', cioè una 'macchina intelligente'. La macchina, attraverso il cosiddetto 'test dell'imitazione', sarebbe stata in grado di 'ingannare' l'operatore, interagendo direttamente con lui attraverso un sistema di comunicazione scritta a distanza e inducendolo a confondere le risposte e i ragionamenti della macchina con quelli di un essere umano (Fasan 2019). La finalità del test era, infatti, di valutare se una macchina fosse in grado di esibire un comportamento 'intelligente', cioè indistinguibile da quello di un essere umano (Marongiu 2020). Turing sostiene che una macchina può essere considerata 'artificialmente intelligente' se è in grado di comprendere il linguaggio naturale e comunicare efficacemente, raccogliere e apprendere informazioni sia da quelle già possedute che dalle interazioni con gli esseri umani, utilizzare i dati archiviati per rispondere a domande e generare nuove soluzioni, e adattarsi infine a circostanze nuove e imprevedibili (Marongiu 2020). Questi studi pionieristici hanno dato impulso a una rapida espansione della ricerca sull'IA a livello globale nei decenni successivi. La formulazione del concetto di Intelligenza artificiale è formalmente attribuita al gruppo di ricerca statunitense guidato da John McCarthy che, nel 1955, per primo considerò l'IA come un nuovo e specifico settore scientifico (Russell e Norvig 2016). Ad oggi, non esiste una concettualizzazione unica e universalmente accettata di IA tra gli esperti del settore (Grosz *et al*. 2016). Tuttavia, tra le definizioni proposte emergono alcuni elementi comuni che permettono di delinearne le caratteristiche principali, evidenziando come l'IA rappresenti un insieme di tecnologie dotate della capacità di apprendere attraverso l'analisi dell'ambiente circostante, eseguire compiti tradizionalmente attribuiti alla creatività e all'ingegnosità umana, migliorare le proprie prestazioni e prendere decisioni autonome per raggiungere obiettivi specifici (Benbya *et al*. 2020; European Commission 2019; Ferilli *et al*. 2021; Susar e Aquaro 2019). In termini di operatività, gli strumenti di IA non sono omogenei e non seguono un'unica logica funzionale. La principale distinzione riguarda la differenza tra IA predittiva e IA generativa. L'IA predittiva utilizza l'apprendimento automatico per identificare schemi ricorrenti in eventi passati e fare previsioni su eventi futuri. L'IA generativa, invece, rende possibile la produzione di nuovi contenuti. Sfruttando tecniche come il *machine learning*, il *deep learning*, l'elaborazione del linguaggio naturale (*natural language processing,* in acronimo NLP) e il riconoscimento vocale, l'IA generativa si riferisce alla creazione di output da parte della macchina, attraverso l'apprendimento e l'analisi dei modelli presenti nei dati di input forniti dall'utente (Aoki 2020; Baldauf e Zimmermann 2020). La principale difficoltà, secondo gli esperti, consiste nel promuovere lo sviluppo di un'IA umano-centrica, cioè orientata all'essere umano.

Conformemente a questa visione, il Rapporto Draghi sulla competitività europea (2024)[2] ha sottolineato l'importanza di investire nello sviluppo di un'IA in grado di favorire l'innovazione tecnologica in piena coerenza con i principi di inclusione sociale. In tale prospettiva, l'IA non va intesa come un sostituto dell'essere umano, bensì come uno strumento progettato per affiancarlo e potenziarne le capacità. A livello europeo, già nelle fasi preparatorie del regolamento sull'Intelligenza artificiale, le istituzioni comunitarie hanno sottolineato l'importanza di integrare le capacità cognitive umane con quelle artificiali, proprie delle macchine, al fine di superare le limitazioni di entrambi i sistemi (European Commission 2019). L'IA rappresenta, quindi, una delle principali frontiere dello sviluppo e dell'innovazione nell'UE. Tuttavia, nonostante le numerose sperimentazioni in corso nel settore privato, l'applicazione di queste tecnologie nel settore pubblico è ancora limitata, impedendo alle amministrazioni e alle istituzioni di sfruttare pienamente i potenziali benefici. (Grimmelikhuijsen e Tangi 2024; van Noordt *et al*. 2023). Sulla base dell'esperienza acquisita attraverso i primi progetti pilota sull'integrazione dell'IA, le amministrazioni pubbliche e le istituzioni

2 Si veda il Report *The future of European competitiveness. A competitiveness strategy for Europe* del settembre 2024, https://commission.europa.eu/topics/eu-competitiveness/draghi-report_en#paragraph_47059.

dell'UE stanno progressivamente affrontando le sfide legate all'implementazione di soluzioni di IA, con l'obiettivo di integrarle in modo sostenibile nei processi organizzativi, nei principi guida e nelle competenze operative di ciascuna organizzazione (Jorge Ricart *et al*. 2022). Dall'analisi critica della letteratura esistente, si propone di seguito una classificazione delle principali opportunità e delle maggiori sfide legate all'integrazione dell'IA generativa nel settore pubblico a livello europeo, accompagnata da esempi di buone pratiche implementate (Aoki 2020; Baldauf e Zimmermann 2020; Berryhill *et al*. 2019; Grimmelikhuijsen e Tangi 2024; Jorge Ricart *et al*. 2022; Medaglia e Tangi 2022; Mergel *et al*. 2024; Mikhaylov *et al*. 2018; OECD e Unesco 2024; Susar e Aquaro 2019; Tangi *et al*. 2023; Tangi *et al.* 2024; van Noordt e Misuraca 2022).

Come delineato nella tabella 1, le principali potenzialità dell'impiego dell'IA generativa nel settore pubblico nell'UE si manifestano in diversi ambiti, tra cui l'ottimizzazione dei processi e delle risorse. In particolare, l'IA consente l'assegnazione delle attività ripetitive alle macchine, con conseguente riduzione dei costi operativi. Inoltre, favorisce l'interoperabilità tra le pubbliche amministrazioni, migliora l'elaborazione di grandi volumi di dati e informazioni, grazie all'utilizzo del *machine learning* che rappresenta una significativa opportunità per l'analisi dei Big Data. L'IA contribuisce anche al miglioramento delle performance nell'erogazione dei servizi pubblici, rispondendo in

Tabella 1. Opportunità dell'integrazione dell'IA generativa nel settore pubblico nell'UE

Opportunità	Descrizione	Best practice UE
Ottimizzazione di processi e risorse	L'automazione di compiti ripetitivi tramite IA consente di ridurre i costi operativi e riallocare le risorse verso attività strategiche. Tra i compiti automatizzati rientrano la classificazione delle richieste dei cittadini e la gestione dei documenti digitali.	Digitalizzazione amministrativa (Germania): l'amministrazione del Baden-Württemberg ha adottato il sistema di assistenza testuale *AI F13* per ottimizzare i processi documentali interni (sintesi, gestione note, ricerca e generazione testi).
Interoperabilità	L'IA promuove l'integrazione e la gestione dei dati tra pubbliche amministrazioni, garantendo interoperabilità a livello legale, organizzativo, semantico e tecnico.	Chatbot unico per i servizi pubblici (Estonia): *Bürokratt* è un chatbot che consente ai cittadini estoni di accedere a tutti i servizi pubblici attraverso un unico strumento, favorendo l'interoperabilità tra diverse amministrazioni e migliorando l'integrazione dei processi a livello nazionale.
Elaborazione di dati e informazioni	Tramite l'elaborazione del linguaggio naturale, l'IA consente l'analisi rapida di documenti amministrativi complessi, come relazioni annuali o dati demografici, facilitando la sintesi per l'elaborazione di politiche pubbliche e riducendo gli errori umani.	Piattaforma di controllo antifrode (Danimarca): la piattaforma *DBA* potenzia l'elaborazione dei dati per individuare irregolarità e frodi, migliorando i controlli e supportando le istituzioni di supervisione finanziaria indipendenti.
Erogazione dei servizi pubblici	L'IA ottimizza la progettazione e l'erogazione dei servizi pubblici, accelerando i tempi di risposta e adottando un approccio focalizzato sulle esigenze del cittadino.	Salute digitale 4.0 (Italia, Ospedale Spallanzani di Roma): durante la pandemia da Covid-19, il servizio *Healthcare Bot* di Microsoft, che usa l'elaborazione del linguaggio naturale, è stato adottato per supportare il triage e rispondere alle FAQ dei cittadini, fornendo informazioni aggiornate e migliorando l'efficienza dei servizi sanitari al cittadino in un contesto di crisi.
Personalizzazione dell'interazione con il cittadino	Attraverso la combinazione dell'IA generativa e di tecniche di analisi dei dati, è possibile offrire un'elevata personalizzazione dell'interazione con il cittadino.	Pianificazione urbana partecipativa (Finlandia): la città di Helsinki ha adottato *UrbanistAI* per coinvolgere attivamente i cittadini, inclusi quelli con disabilità, nella co-creazione di soluzioni urbane personalizzate tramite IA generativa.

Fonte: elaborazione dell'Autrice

modo più efficace alle esigenze dei cittadini, e alla personalizzazione dell'interazione con gli utenti. Le principali criticità riguardano l'attuale capacità delle infrastrutture digitali degli enti pubblici di supportare l'effettiva implementazione dell'IA, le potenziali distorsioni algoritmiche nell'utilizzo di tali strumenti, le resistenze culturali interne e la conseguente inadeguatezza del personale pubblico in termini di formazione e competenze. Questo richiede l'implementazione di corsi di formazione specifici, con un ulteriore impatto sul bilancio pubblico. A ciò si aggiungono le problematiche relative alla privacy e alla protezione dei dati, che sollevano importanti interrogativi etici sulla gestione delle informazioni personali dei cittadini-utenti (tabella 2).

Nel complesso, riflettere congiuntamente sulle opportunità e sulle sfide legate all'integrazione dell'IA generativa nel settore pubblico è essenziale per promuovere un approccio umano-centrico, incentrato sui cittadini e orientato al bene comune. Se opportunamente gestiti, gli strumenti di IA possiedono il potenziale per contribuire significativamente a tale obiettivo, fungendo da catalizzatori di quella che può essere definita 'intelligenza istituzionale' (Marongiu 2020). Se non propriamente gestiti possono determinare, invece, la condizione opposta di 'opacità istituzionale' (Janssen *et al*. 2020). Sebbene sia prematuro trarre conclusioni definitive sull'efficacia dell'IA generativa nel contesto del settore pubblico europeo, la recente indagine condotta dall'Osservatorio europeo Public Sector Tech Watch offre una panoramica rilevante ai fini di questo contributo (Grimmelikhuijsen e Tangi 2024). Con l'obiettivo di raccogliere informazioni sulla percezione e sull'impiego attuale dell'IA generativa da parte delle organizzazioni del settore pubblico nell'UE, lo studio ha coinvolto 579 dirigenti pubblici provenienti da sette Stati membri (*ibidem*). Il 30% degli intervistati ha dichiarato di aver già integrato strumenti di questo tipo a supporto delle proprie attività professionali, mentre il 44% prevede di adottarli in futuro. Il 26% restante ha, invece,

Tabella 2. Sfide dell'integrazione dell'IA generativa nel settore pubblico nell'UE

Sfida	Descrizione	Soluzione proposta
Inadeguatezza delle infrastrutture digitali	Le infrastrutture digitali esistenti, come le piattaforme cloud, spesso non sono abbastanza robuste, scalabili o sicure per supportare efficacemente l'implementazione dell'IA nel settore pubblico.	Investire in infrastrutture digitali scalabili e interoperabili, con supporto finanziario e tecnico dell'UE.
Bias algoritmici	L'IA può fornire risposte imprecise e non del tutto trasparenti, soprattutto se non è stata adeguatamente programmata o se non ha accesso a insiemi di dati (dataset) di informazioni aggiornati.	Garantire dataset inclusivi e rappresentativi, implementando audit periodici e linee guida etiche per gli sviluppatori.
Inerzia istituzionale	Il personale può opporsi all'adozione dell'IA per timori di sostituzione del lavoro umano, preferendo mantenere pratiche tradizionali e resistendo al cambiamento.	Introdurre programmi di sensibilizzazione ad hoc all'interno di istituzioni e pubbliche amministrazioni sull'utilizzo etico e umano-centrico dell'IA nel settore pubblico.
Competenze e professionalizzazione	Il personale pubblico presenta carenze nelle competenze necessarie per utilizzare l'IA a supporto del proprio lavoro. Questa situazione comporta una carenza di profili professionali esperti, capaci di garantire un'integrazione efficace di questi strumenti a complemento delle attività umane.	Promuovere corsi di formazione mirati allo sviluppo di competenze specifiche sull'uso dell'IA nel settore pubblico. Questi programmi dovrebbero includere moduli pratici sull'analisi dei dati, l'elaborazione del linguaggio naturale e l'automazione dei processi amministrativi, affiancati da aggiornamenti periodici per garantire un costante adeguamento alle innovazioni tecnologiche.
Privacy e protezione dei dati personali	L'uso dell'IA nella gestione dei dati dei cittadini comporta il rischio di violazioni della privacy e della sicurezza, in particolare quando vengono trattate informazioni sensibili.	Applicare standard di sicurezza avanzati, come la crittografia end-to-end, e promuovere l'uso di IA open source all'interno dell'UE.

Fonte: elaborazione dell'Autrice

espresso una certa reticenza nell'adottare queste tecnologie. Tra coloro che non utilizzano questi strumenti (77%), i principali ostacoli individuati sono la mancanza di conoscenza (41%), di fiducia (25%) e di consapevolezza (11%). Le preoccupazioni etiche, come quelle relative alla condivisione dei dati (22%) e all'impatto ambientale (4%), sono state segnalate con minore frequenza. Solo il 16% degli intervistati ha giudicato questi strumenti come inutili. Tra le funzioni più utilizzate, l'ottimizzazione dei testi (92%) e quella dei sistemi di domanda-risposta Q&A (90%). Il beneficio principale percepito è rappresentato dall'efficienza, intesa come riduzione dei tempi (77%), seguito dallo sviluppo di nuove competenze (74%). Le aspettative relative al supporto dei processi decisionali e al miglioramento della comunicazione attraverso l'uso dell'IA sono ritenute inferiori (54%). Nel complesso, emerge un sentiment positivo generalizzato che evidenzia la necessità di un monitoraggio continuo del fenomeno, con l'opportunità di estendere l'analisi all'intero contesto europeo e di ampliare il campione includendo gli operatori con funzioni non dirigenziali.

3. Chatbot istituzionali: la nuova frontiera della comunicazione pubblica europea?

In una UE sempre più digitale, l'uso dell'IA generativa consente alle organizzazioni pubbliche di migliorare l'efficienza dei servizi e di offrire un'esperienza utente più personalizzata (Androutsopoulou *et al*. 2019; Lovari e De Rosa 2025; Ng *et al*. 2021; Pane 2025). Queste tecnologie innovative trovano applicazione anche nell'ambito della comunicazione pubblica, sia per le attività interne che per quelle rivolte all'esterno. I chatbot, noti anche come assistenti virtuali (*virtual assistants*) o interfacce conversazionali (*conversational AI*), sono applicazioni software progettate per interagire con gli utenti attraverso il linguaggio naturale, sia in forma testuale che vocale. Grazie all'elaborazione del linguaggio naturale, che consente una risposta rapida e personalizzata alle richieste dell'utente, i chatbot offrono supporto in tempo reale e sono disponibili 24 ore su 24, garantendo un'assistenza continua e immediata (Androutsopoulou *et al*. 2019; Aoki 2020; Baldauf e Zimmermann 2020; Berryhill *et al*. 2019; Susar e Aquaro 2019; van Noordt e Misuraca 2022). Sempre dal punto di vista dell'utente, i chatbot alleviano il sovraccarico informativo presentando le informazioni più rilevanti per il servizio richiesto (van Noordt e Misuraca 2022). Secondo Makasi *et al*. (2022), si distinguono in due principali categorie: chatbot di base e chatbot avanzati. I primi si avvalgono di modelli algoritmici orientati al recupero delle informazioni, presentano capacità limitate di elaborazione del linguaggio naturale e utilizzano i dati in tempo reale in misura ridotta (Makasi *et al*. 2022). I secondi, invece, adottano modelli algoritmici potenziati, sono dotati di funzionalità più sofisticate di elaborazione del linguaggio naturale e sfruttano pienamente i dati in tempo reale per offrire servizi più dinamici e personalizzati (*ibidem*). Androutsopoulou *et al*. (2019) mettono in evidenza le potenzialità dei chatbot avanzati che, nel fornire un servizio all'utente, sono in grado di analizzare le richieste e selezionare le opzioni più adeguate. Inoltre, i chatbot, nel fornire il loro servizio secondo la logica input-output precedentemente descritta, si basano su dataset preesistenti che, se incompleti o non aggiornati, possono generare risposte imprecise, compromettendo l'efficacia della comunicazione pubblica (Aoki 2020; Baldauf e Zimmermann 2020). Ad esempio, un chatbot che utilizza un dataset obsoleto potrebbe fornire informazioni errate su orari o requisiti amministrativi, generando confusione tra i cittadini e compromettendo la fiducia nelle istituzioni. Un'ulteriore sfida consiste nel garantire che la comunicazione pubblica europea, nel nuovo paradigma '4.0', riesca a bilanciare in modo equilibrato il ruolo del cittadino-utente, degli algoritmi e delle piattaforme digitali. È fondamentale evitare che la dimensione tecnologica prenda il sopravvento su quella umana, riducendo così il rischio di una comunicazione istituzionale deumanizzata (Kim e McGill 2024). Per evitare tale rischio, la progettazione dei chatbot istituzionali deve mantenere al centro l'esperienza del cittadino, utilizzando linguaggi semplici e accessibili che simulino un'interazione umana. A questo fine, l'interfaccia dei chatbot istituzionali è stata progettata per simularla, semplificando il dialogo con la macchina grazie all'uso di un linguaggio semplice e informale, in alternativa al registro formale tipico delle istituzioni (Makasi *et al*. 2020). Negli ambienti europei, l'uso dei chatbot istituzionali – come strumento operativo di supporto alla comunicazione pubblica – è attualmente in fase di sperimentazione. Con riferimento alla comunicazione interna, un esempio di applicazione è rappresentato da *GPT@EC*, lo strumento sviluppato dalla Direzione generale per i Servizi digitali della Commissione

europea, basato sul precedente progetto di successo di *GPT@JRC*, implementato dal Centro comune di ricerca dell'Esecutivo europeo per scopi scientifici e di ricerca nell'ambito dell'IA generativa. Sul piano operativo, *GPT@EC* ha l'obiettivo di migliorare la produttività e la creatività del personale, supportandolo in attività quali la redazione di testi, la sintesi di documenti e lo sviluppo di codici software, utilizzando modelli linguistici di grandi dimensioni (*Large language models*) open source. Lo strumento costituisce un'alternativa alle soluzioni di IA generativa di terze parti, riducendo i rischi legati a una gestione impropria dei dati sensibili della Commissione. In particolare, questa soluzione permette di elaborare le informazioni interne senza doverle condividere con soggetti esterni. Per quanto riguarda, invece, la comunicazione pubblica europea esterna, cioè rivolta al cittadino-utente, un esempio significativo di chatbot istituzionale è l'*Eurodesk Mobility Advisor Chatbot*. Lo strumento è configurato con quattro funzionalità che consentono di personalizzare l'interazione con il cittadino. La prima funzione *EMA* (*Eurodesk Mobility Advisor*) fornisce informazioni sulle principali iniziative europee rivolte ai giovani (Eurodesk, Erasmus+, Corpo europeo di solidarietà e DiscoverEU) attraverso un linguaggio adatto a un pubblico generalista e variegato. La seconda funzione, *Snoop Dogg*, offre un servizio simile, ma predilige un approccio linguistico informale, specificamente pensato per un target più giovanile. La terza funzione, *Altydesk*, consente di analizzare un'immagine caricata dal cittadino sulla piattaforma, identificando le informazioni più rilevanti e trascrivendole in formato testuale. Infine, la quarta funzione, *TLDR* (*Too Long; Didn't Read*), permette di riassumere testi più lunghi caricati dal cittadino-utente in piattaforma. Oltre alle funzionalità tecniche, un aspetto cruciale per il successo di questi strumenti risiede nella cura del design e nella centralità delle esigenze dei cittadini. Riflettere sull'impiego dell'IA a supporto del settore pubblico, e in particolare nell'erogazione dei servizi pubblici, secondo Reale (2023), non può prescindere da un'analisi approfondita del design di tali strumenti, che deve essere progettato tenendo conto delle esigenze dei cittadini Nel caso dell'*Eurodesk Mobility Advisor Chatbot*, l'interfaccia del servizio è disponibile in tutte le 24 lingue ufficiali dell'UE e offre una serie di impostazioni personalizzabili, consentendo all'utente di contribuire, seppur in maniera limitata, al design del servizio. Il cittadino ha, infatti, la possibilità di personalizzare vari aspetti del chatbot, come la scelta dei colori e dei caratteri, adattando così l'interfaccia alle proprie preferenze o necessità. In definitiva, l'adozione di chatbot istituzionali e di strumenti basati su IA generativa rappresenta un'opportunità significativa, ma anche una sfida epocale, per migliorare la comunicazione pubblica e l'erogazione dei servizi nel settore pubblico europeo. Questo contesto richiede un monitoraggio costante a medio-lungo termine, oltre a un profondo impegno istituzionale, per valutare l'impatto effettivo di tale strumento sulla comunicazione pubblica dell'UE.

Conclusioni

L'adozione dell'IA generativa nel settore pubblico rappresenta una frontiera ricca di opportunità, ma anche di sfide significative che richiedono una riflessione approfondita e multidisciplinare. Nell'UE, le organizzazioni hanno avviato le prime sperimentazioni di queste tecnologie, con un focus particolare sulla comunicazione pubblica. La transizione digitale ha infatti spinto l'evoluzione della comunicazione pubblica verso un nuovo paradigma, noto come 'comunicazione pubblica 4.0', in cui il cittadino, inteso come utente, condivide il ruolo di protagonista con gli algoritmi e, più in generale, con le piattaforme digitali. In questo scenario innovativo, sono stati introdotti i primi chatbot istituzionali, progettati sia per migliorare la comunicazione interna delle amministrazioni e delle istituzioni, sia per ottimizzare quella esterna rivolta ai cittadini. L'obiettivo principale di questi strumenti è duplice: da un lato, semplificare e rendere più efficienti i processi interni, e dall'altro, personalizzare l'interazione con i cittadini, favorendo così un dialogo più diretto e accessibile. Tali iniziative rappresentano un importante banco di prova per esplorare le potenzialità dell'IA generativa nel migliorare l'efficienza dei processi comunicativi e promuovere un dialogo bidirezionale più inclusivo, determinando un potenziale miglioramento nell'erogazione dei servizi pubblici. Tuttavia, permangono interrogativi cruciali riguardo alle implicazioni etiche, in particolare per quanto concerne la sicurezza dei dati, la distorsione algoritmica e la trasparenza amministrativa, oltre a quella processuale, e dei meccanismi di funzionamento delle 'macchine intelligenti'. Opportunità da un lato, sfide dall'altro. Per garantire un bilanciamento adeguato lungo

questo continuum, l'UE ha adottato un modello di sviluppo dell'IA umano-centrico, orientato al benessere e alle esigenze dell'essere umano. A tale scopo, l'UE si è dotata del primo quadro normativo per l'IA, che, attualmente, non ha equivalenti a livello globale. Questo quadro mira a garantire trasparenza, sicurezza e inclusività nell'adozione e nell'utilizzo di queste tecnologie emergenti. Certamente, è ancora presto per stabilire se le opportunità offerte dall'IA generativa saranno pienamente sfruttate, se le sfide attualmente individuate saranno affrontate in modo efficace e se emergeranno nuove criticità. Ciò che appare certo è che la comunicazione pubblica in Europa, così come le funzioni del settore pubblico in generale, si trovano di fronte a un importante banco di prova, volto a testare la resilienza e la capacità di cogliere le opportunità offerte dal progresso tecnologico. A tal fine, sarà fondamentale monitorare l'utilizzo degli strumenti di IA generativa e promuovere lo sviluppo di competenze critiche tra i dipendenti e i funzionari pubblici, al fine di garantire un impiego etico e consapevole di queste tecnologie. Solo attraverso un'attenzione costante a questi elementi sarà possibile garantire lo sviluppo umano-centrico dell'IA, inteso come una prerogativa fondamentale per il raggiungimento della condizione di 'intelligenza istituzionale', considerata un asset strategico per rispondere in modo più efficace e tempestivo alle sfide di un'Unione in continua evoluzione.

Bibliografia

Androutsopoulou A., Karacapilidis N., Loukis E., Charalabidis Y. (2019), Transforming the communication between citizens and government through AI-guided chatbots, *Government information quarterly*, 36, n.2, pp.358-367

Aoki N. (2020), An experimental study of public trust in AI chatbots in the public sector, *Government information quarterly*, 37, n.4 <DOI:10.1016/j.giq.2020.101490>

Baldauf M., Zimmermann H.D. (2020), Towards conversational E-government: An experts'perspective on requirements and opportunities of voice-based citizen services, in *HCI in Business, Government and Organizations: 7th International Conference, HCIBGO 2020, Held as Part of the 22nd HCI International Conference, HCII 2020*, Copenhagen, Denmark, July 19-24, Berlin, Springer-Verlag, pp.3-14

Benbya H., Davenport T.H., Pachidi S. (2020), Artificial intelligence in organizations: Current state and future opportunities, *MIS Quarterly Executive*, 19, n.4 <DOI: 10.2139/ssrn.3741983>

Berryhill J., Heang K.K., Clogher R., McBride K. (2019), *Hello, World. Artificial intelligence and its use in the public sector*, OECD Working Papers on Public Governance n.36, Paris, OECD Publishing

Canel M.J., Luoma-Aho V. (2018), *Public sector communication, Closing gaps between citizens and public organizations*, Hoboken, John Wiley & Sons

Canel M.J., Sanders K. (2013), Introduction: Mapping the field of government communication, in Canel M.J., Sanders K. (eds.), *Government communication. Cases and challenges*, London, Bloomsbury Academy, pp.1-26

D'Ambrosi L. (2019), *La comunicazione pubblica dell'Europa. Istituzioni, cittadini e media digitali*, Roma, Carocci

Ducci G. (2013), La comunicazione pubblica digitale dell'Europa, per l'Europa, *Sociologia della Comunicazione*, 44, pp.79-97

European Commission (2019), *Ethics guidelines for trustworthy AI*, Brussell, European Commission

Fasan M. (2019), Intelligenza Artificiale e pluralismo: uso delle tecniche di profilazione nello spazio pubblico democratico, *BioLaw Journal-Rivista di BioDiritto*, 1, pp.101-113

Ferilli S., Emanuela G., Musto C., Marina P., Piero P., Silvia P., Semeraro G. (2021), *L'Intelligenza Artificiale per lo Sviluppo Sostenibile*, Roma, CNR Edizioni

Floridi L. (2021), *Ethics, governance, and policies in artificial intelligence*, Cham, Springer

Floridi L. (2020), Artificial intelligence as a public service. Learning from Amsterdam and Helsinki, *Philosophy & Technology*, 33, n.4, pp.541-546

Grimmelikhuijsen S., Tangi L. (2024), *What factors influence perceived artificial intelligence adoption by public managers? A survey among public managers in seven EU countries*, Luxembourg, Publications Office of the European Union

Grosz B.J., Mackworth A., Altman R., Horvitz E., Mitchell T., Mulligan D., Shoham, Y. (2016), *Artificial intelligence and life in 2030: One hundred years study on artificial intelligence*, Stanford, Stanford University

Janssen M., Hartog M., Matheus R., Yi Ding A., Kuk G. (2020), Will algorithms blind people? The effect of explainable AI and decision-makers'experience on AI-supported decision-making in government, *Social Science Computer Review*, 40, n.1, pp.478-493

Jorge Ricart R., Van Roy V., Rossetti F., Tangi L. (2022), *AI Watch – National strategies on artificial intelligence. A European perspective*, Luxembourg, Publications Office of the European Union

Kim H.Y., McGill A.L. (2024), AI-induced dehumanization, *Journal of Consumer Psychology*, 35, n.3, pp.363-381

Laux J., Wachter S., Mittelstadt B. (2024), Trustworthy artificial intelligence and the European Union AI Act. On the conflation of trustworthiness and acceptability of risk, *Regulation & Governance*, 18, n.1, pp.3-32

Lovari A., Belluati M. (2023), We Are All Europeans. EU Institutions Facing the Covid-19 Pandemic and Information Crisis, in La Rocca G., Carignan M.E., Artieri G.B. (eds.), *Infodemic Disorder: Covid-19 Coping Strategies in Europe, Canada and Mexico*, Cham, Springer International Publishing, pp.65-96

Lovari A., De Rosa F. (2025), Exploring the Challenges of Generative AI on Public Sector Communication in Europe, *Media and Communication*, 13 <DOI:10.17645/mac.9644>

Lovari A., Ducci G. (2022), *Comunicazione pubblica*, Milano, Mondadori Università

Lovari A., D'Ambrosi L., Bowen S. (2020), Re-connecting voices. The (new) strategic role of public sector communication after Covid-19 crisis, *Partecipazione e conflitto*, 13, n.2, pp. 970-989

Macnamara J. (2017), *Evaluating public communication: Exploring new models, standards, and best practice*, London, Routledge

Makasi T., Nili A. Desouza K., Tate M. (2020), Chatbot-mediated public service delivery: A public service value-based framework, *First Monday*, 25, n.12 <DOI:10.5210/fm.v25i12.10598>

Manzella A. (2000), La ripartizione di competenze tra Unione europea e Stati membri, *Quaderni costituzionali*, 20, n.3, pp.531-544

Marongiu D. (2020), L'Intelligenza Artificiale "istituzionale": limiti (attuali) e potenzialità, *European Review of Digital Administration & Law*, 1, pp.37-54

McCulloch W.S., Pitts W. (1943), A logical calculus of the ideas immanent in nervous activity, *The bulletin of mathematical biophysics*, 5, n.4, pp.115-133

Medaglia R., Tangi L. (2022), The adoption of Artificial Intelligence in the public sector in Europe: drivers, features, and impacts, in *Proceedings of the 15th International Conference on Theory and Practice of Electronic Governance*, New York, Association for Computing Machinery, pp.10-18

Mergel I., Dickinson H., Stenvall J., Gasco M. (2024), Implementing AI in the public sector, *Public Management Review*, pp.1-14 <DOI:10.1080/14719037.2023.2231950>

Mikhaylov S.J., Esteve M., Campion A. (2018), Artificial intelligence for the public sector: opportunities and challenges of cross-sector collaboration, *Philosophical transactions of the royal society a: mathematical, physical and engineering sciences*, 376, n.2128, 20170357<DOI:10.1098/rsta.2017.0357>

Mökander J., Axente M., Casolari F., Floridi L. (2022), Conformity assessments and post-market monitoring: A guide to the role of auditing in the proposed European AI regulation, *Minds and Machines*, 32, n.2, pp.241-268

Nieto A. (2006), *Economia della comunicazione istituzionale,* Milano, Franco Angeli

Ng D.T.K., Leung J.K.L., Chu S.K.W., Qiao M.S. (2021), Conceptualizing AI literacy: An exploratory review, *Computers and Education: Artificial Intelligence*, 2, n.2008, 100041 <DOI:10.1016/j.caeai.2021.100041>

OECD (2021), *OECD Report on Public Communication: The Global Context and the Way Forward*, Paris, OECD Publishing

OECD, Unesco (2024), *G7 Toolkit for Artificial Intelligence in the Public Sector*, Paris, OECD Publishing

Pane S. (2025), The european 'post-digital'public sphere: foundations of an emerging paradigm in the social sciences, *Methaodos. revista de ciencias sociales*, 13, n.1, <DOI:10.17502/mrcs.v13i1.866>

Pollitt C., Van Thiel S., Homburg V. (eds.) (2007), *New Public Management in Europe: adaptations and alternatives*, Basingstoke, Palgrave MacMillan

Pollitt C, Dan S. (2011), *The impacts of the New Public Management in Europe: A meta-analysis*, Cocops work package 1, Rotterdam, CoCoPS – Coordinating for Cohesion in the Public Sector of the Future

Reale R. (2023), L'AI generativa e i servizi pubblici digitali: scenario e possibili applicazioni, *Forumpa.it*, 5 maggio

Parlamento europeo, Consiglio europeo (2024), *Regolamento 2024/1689 del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'Intelligenza Artificiale e modifica i regolamenti (CE) 2008/300, 167/2013,*

2013/168, 2018/858, 29018/1139 e 2019/2144 e le direttive 2014/90, 2016/797 2020/1828 (regolamento sull'Intelligenza Artificiale), in GU Serie L del 12.7.2024

Russell S.J., Norvig P. (2016), *Artificial intelligence: a modern approach*, London, Pearson

Smillie L., Scharfbillig M. (2024), *Trustworthy Public Communications*, Luxembourg, Publications Office of the European Union

Susar D., Aquaro V. (2019), Artificial intelligence: Opportunities and challenges for the public sector, *Proceedings of the 12th international conference on theory and practice of electronic governance*, New York, Association for Computing Machinery, pp.418-426

Tangi L., Combetto M., Hupont Torres I., Farrell E., Schade S. (2024), *The potential of generative AI for the public sector: Current use, key questions and policy considerations*, Bruxelles, European Commission, Joint Research Centre

Tangi L., Combetto M., Martin B.J., Rodriguez M.P. (2023), *Artificial Intelligence for Interoperability in the European Public Sector*, Luxembourg, Publications Office of the European Union

Turing A. (1950), Computing machinery and intelligence, *Mind*, 59, n.236, pp.433-460

van Noordt C., Misuraca G. (2022), Artificial intelligence for the public sector: results of landscaping the use of AI in government across the European Union, *Government information quarterly*, 39, n.6, 101714 <DOI:10.1016/j.giq.2022.101714>

van Noordt C., Medaglia R., Tangi L. (2023), Policy initiatives for Artificial Intelligence-enabled government: An analysis of national strategies in Europe, *Public Policy and Administration*, 40, n.2, pp.215-253

Zémor P. (1995), *La communication publique*, Paris, Presses Universitaires de France

Sara Pane
sara.pane@uniroma1.it
Dottoranda PNRR presso l'Università Sapienza di Roma, collabora con Europe Direct Città Metropolitana di Torino ed è teaching staff e coordinatrice della comunicazione per la Jean Monnet Chair Com4T.EU presso l'Università di Torino. Tra le pubblicazioni recenti si segnala: The European 'post-digital'public sphere: Foundations of an emerging paradigm in the social sciences. Methaodos, *Revista de Ciencias Sociales*, n.1, 2025.

An examination of fundamental rights protection in AI governance

Laura Evangelista
INAPP

Concetta Fonzo
INAPP

Eleonora Zecca
LUISS

This article provides an analysis of the impact of Artificial intelligence (AI) on contemporary society, highlighting the numerous violations of fundamental rights, including the rights to privacy, non-discrimination, and access to information. Drawing on concrete examples and in light of the inadequacy of current legislation, the paper examines existing jurisprudence on AI, beginning with the *Artificial intelligence Act* adopted by the European Commission in 2024 and comparing it with U.S. legislation to evaluate differing regulatory approaches. The article concludes by discussing the future challenges involved in developing a trustworthy and reliable AI system capable of safeguarding fundamental rights and ensuring algorithmic accountability.

L'articolo offre un'analisi dell'impatto che l'Intelligenza artificiale (IA) ha sulla società odierna, nonché delle numerose violazioni di diritti fondamentali, come il diritto alla privacy, alla non discriminazione e all'informazione. Alla luce di esempi concreti e della legislazione attuale insufficiente, il contributo esamina la giurisprudenza esistente sull'IA, a partire dall'Artificial intelligence Act adottato dalla Commissione europea nel 2024, confrontandolo con la legislazione degli Stati Uniti per valutare i diversi approcci. Il lavoro termina con una disamina delle sfide future verso un sistema di IA affidabile, che garantisca la protezione dei diritti e algoritmi attendibili.

DOI: 10.53223/Sinappsi_2025-02-6

Citazione
Evangelista L., Fonzo C., Zecca E. (2025), An examination of fundamental rights protection in AI governance, *Sinappsi,* XV, n.2, pp.75-89

Parole chiave
Artificial intelligence
European policies
Fundamental rights

Keywords
Intelligenza artificiale
Politiche comunitarie
Diritti fondamentali

Introduction

In just a few decades, Artificial intelligence (AI) has shifted from the realm of science fiction to a significant force influencing nearly every facet of contemporary life, redefining the limits of machine capabilities. Central to this technological transformation is a paradox: AI possesses the potential to significantly improve human lives in ways never seen before, yet it also introduces substantial risks related to privacy, misinformation, and discrimination. To date, scholars refer to AI as the development of computer systems designed to perform tasks that typically require human intelligence, such as visual perception, speech recognition and decision-making[1], although a

1 On this point, for instance, see: Copeland B.J., Artificial intelligence, in *Encyclopedia Britannica,* available at: https://www.britannica.com/technology/artificial-intelligence.

universally accepted definition of artificial intelligence (West 2018) still seems to be lacking.

Moreover, while many might assume that studies on AI began recently, Alan Turing is often recognised as the pioneer of the concept. In 1950, he theorised about "thinking machines" capable of reasoning like humans. His "Turing Test" stated that computers must complete reasoning tasks at a level comparable to humans in order to be considered "thinking" in an independent manner (Turing 1950). Later, building on Turing's ideas, McCarthy was the first to introduce the term "Artificial intelligence" to describe the field focused on creating intelligent machines, particularly software. According to McCarthy, this field is linked to the challenge of employing computers to understand human intelligence (McCarthy 2022). The issue, however, is that it's still impossible to define what types of computational processes can be classified as intelligent, as the mechanisms underlying human intelligence are not yet fully understood (*ibidem*). These challenges became apparent with the development of the first AI machines. The early wave of Artificial intelligence proposed a "Symbolic AI" system which aimed to create computers that could mimic the problem-solving skills of the human brain[2]. A key application of "Symbolic AI" was developed at the Massachusetts Institute of Technology in 1966 by Joseph Weizembaum, who developed *ELIZA,* a chatbot programme designed to simulate conversations between a patient and a psychotherapist (Treusch 2020). *ELIZA* was based on a script containing rules, and input sentences were restricted to simple declarative and interrogative forms based on known objects and relationships. In order to communicate, *ELIZA* used patterns and predefined rules to generate responses (*ibidem*). Despite general enthusiasm for Artificial intelligence, the "Symbolic AI" approach ultimately failed. It became clear that many fields or domains — language, for instance — cannot be precisely formalised through rules and symbolic manipulation. This first attempt at AI faced significant criticism. Hubert Dreyfus (1972) argued that "Symbolic AI" failed because it was based on philosophical assumptions that compared the brain to a computer (Kenaw 2008). He argued that the brain does not operate like a computer executing distinct tasks, nor does the mind process information solely based on formal principles. Dreyfus also questioned the possibility that all knowledge could be completely articulated in logical or mathematical formats and that reality could be broken down into discrete facts represented by symbols (*ibidem*).

After an "AI winter", a renewed effort to advance the field emerged with the "Connectionist AI", which theorised that machines should simulate neural networks[3] (Hardesty 2017), inspired by the brain structures and functions, rather than relying on human reasoning and logic, as "Symbolic AI" had unsuccessfully attempted (Smolensky 1987). This new approach was more promising, as it was data-driven and capable of learning from examples rather than being rule-based. Following this approach, Frank Rosenblatt developed *The Perceptron* (1958), one of the earliest artificial neural networks. It used a system of interconnected units (like biological neurons) to adaptively classify inputs (Block 1962). This early application of "Connectionist AI" aimed to represent intelligence through networks of simple processing units. In contrast to "Symbolic AI", which employed explicit rules, "the Perceptron" relied on probabilistic models, allowing it to learn and adjust by forming new associations among its components (*ibidem*). However, the main limitation of Perceptron's single-layer structure was in identifying only linearly separable patterns, a limitation later overcome by the development of multi-layer neural networks. Geoffrey Hinton, widely regarded as the "Godfather of Deep Learning", played a crucial role in advancing neural network research. He was able to develop unsupervised pre-training for deep networks, which proved instrumental in maintaining progress in the field and laid the foundation for modern deep learning (LeCun *et al.* 2015). Through his works, Hinton could demonstrate the power of multi-layer

2 For an overview of symbolic AI's evolution, from its philosophical foundations in the 1950s to its modern-day fusion with neural networks, see: *The Evolution of Symbolic AI: From Early Concepts to Modern Applications*, available at: https://smythos.com/developers/agent-development/history-of-symbolic-ai/.

3 Definition: "neural nets are a means of doing machine learning, in which a computer learns to perform some tasks by analysing training examples. Usually, the examples have been hand-labelled in advance". Retrieved from: https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414.

neural networks[4] (deep neural networks), and their ability to learn complex, non-linear patterns and derive hierarchical features directly from raw data. In recognition of his contributions, he was awarded the Nobel Prize in 2024.

Over the years, AI has made significant progress, achieving the appearance of intelligent behaviour by operating through algorithms[5] and statistical learning techniques on extensive datasets (Neapolitan 2012). This made possible concrete advancements, such as the development of "Large Language Models (LLMs)", which generate text that mimics human writing and supports various tasks (Blank 2023).

AI rapid advancement, however, poses significant challenges to fundamental rights. AI-driven systems increasingly threaten privacy and contribute to misinformation and discrimination. Large Language Models often amplify biases already present in their data and propagate harmful stereotypes and hegemonic narratives, despite lacking true understanding of language or possessing real-world knowledge, leading to potential misuse. This occurs because LLMs are pre-trained using massive amounts of data coming from publicly available internet sources, thus often providing low-quality and unrepresentative data, contributing to the spread of misinformation. This has prompted concerns about potential harmful data colonialism by big AI companies, which can easily extract human life data, tracking consumers behaviours, targeting them with personal-branded online content (Couldry and Mejias 2019). This form of mass surveillance contributes to spreading misinformation through AI-generated content, undermining public trust and democratic processes, reinforcing discrimination, amplifying biases and inequalities.

Starting from such a scenario, this article aims to reflect on the impact of artificial intelligence on society, focusing on how its adoption may harm fundamental rights, such as the right to privacy, fair access to information, and non-discrimination, providing concrete case studies as application of such violations in the use of AI tools. This contribution will explore and examine how the newly adopted European Artificial Intelligent Act (European Union 2024, hereinafter AI Act) seeks to protect individuals from the disproportionate use of AI tools, comparing it with legislative tools adopted in the United States, in order to evaluate different legislative approaches and measures. Finally, this article will outline future challenges and goals for improving the development of Artificial intelligence, aiming for a trustworthy AI system that ensures rights protection and reliable safe algorithmic functions. Nevertheless, this challenge is still at the very beginning. At present, it is not possible to develop robust AI systems capable of operating reliably under all conditions, therefore human oversight remains essential.

1. Artificial intelligence and rights: opportunities and threats

Even though AI technologies hold the promise of improving the protection of human rights and the rule of law, they also pose significant risks, potentially leading to discrimination, gender inequality, threats to democratic process, and violations of human rights (Council of Europe 2024). Firmly aware of these threats, the European Union has been committed to ensuring that the development of AI aligns with established human rights principles. In line with this commitment, the Steering Committee for Human Rights (CDDH) has been tasked by the Committee of Ministers with focusing on this emerging area, establishing the Drafting Group on Human Rights and Artificial intelligence (CDDH-IA), which is responsible for creating a Handbook on Human Rights and Artificial intelligence by the end of 2025[6].

Examining the influence of AI on human rights is complex. Many documents primarily focus on identifying the rights and freedoms that could be affected, rather than actively evaluating this potential impact and suggesting assessment frameworks (Mantelero 2022). For this reason, the

4 Definition: "a multilayer neural network has multiple layers stacked on top of each other. Each layer receives input from the previous layer and applies a mathematical operation called an activation function, such as the sigmoid function. This allows the network to capture complex relationships between inputs and outputs". Retrieved from: ttps://muneebsa.medium.com/deep-learning-101-lesson-9-multi-layer-neural-network-7cd3a53066c8.

5 Definition: "a set of mathematical instructions or rules that, especially if given to a computer, will help to calculate an answer to a problem". Retrieved from: https://dictionary.cambridge.org/dictionary/english/algorithm.

6 On this point, see: Council of Europe, *Artificial intelligence and Human Rights (CDDH-IA)*, available at: https://www.coe.int/en/web/human-rights-intergovernmental-cooperation/intelligence-artificielle.

present article analyses the risks AI systems pose to human rights from a legal perspective, setting aside ethical concerns, while providing a case study for each type of rights violation to better understand real potential implications and consequences.

Artificial intelligence and the right to information

The rise of new Generative AI (GAI) technologies[7] can manipulate and create several types of digital content following specific guidelines easily enabling multi-modal misinformation (Rombach 2022). For instance, the increasing amount of Artificial intelligence Generated Content (AIGC) hinders the battle against misinformation (Sohail *et al.* 2023). However, within the realm of fake news, it is necessary to differentiate between disinformation and misinformation. Disinformation is defined as "false, inaccurate or misleading information" spread to deceive the recipients, whereas misinformation "refers to false, inaccurate or misleading information that is shared without any intent to deceive" (Bontridder and Poullet 2021). When AI is involved in the creation of false or fake content, the result is known as a "deepfake"[8]. More specifically, deepfakes are the result of two AI algorithms operating in a so-called Generative Adversarial Network (GAN), a method for algorithmically generating new data from existing datasets (*ibidem*). AI-driven personalisation algorithms have emerged as powerful yet problematic amplifiers of misinformation across social media platforms. This is due to the economic model adopted by the major platforms, known as "the economics of attention" (Festré 2015), whereby investments are indirectly financed through revenues from advertising firms (Bontridder and Poullet 2021). Consequently, the more time users spend on a platform, the more that platform profits (*ibidem*). Users, by interacting with the platform, inform the algorithm about their preferences, what they like, disagree with, think or feel. As a result, the algorithms recommend content that closely aligns with the users interests and opinions, which drives greater profits but, at the same time, arms the user, who ends up seeing and reading only one perspective. In addition, the echo chambers and filter bubbles play a pivotal role in the spread of misinformation and disinformation. The former, often created as a consequence of the latter, refer to contexts in which people are exposed exclusively to viewpoints that reinforce their own beliefs, disregarding opposing perspectives (Festré 2015). This is exactly what occurs with AI-driven personalisation algorithms. Another factor to consider in this discussion is Generative AI, such as ChatGPT and Stable Diffusion, which are increasingly popular, sophisticated algorithms capable of producing highly realistic and convincing content across various forms, including text, images, audio, and video (Seow *et al.* 2022). A key example of Generative AI application has been the spread of fake robocalls from Joe Biden in New Hampshire (Hsu 2024). Using Generative AI, a company had published fake recordings of US President Joe Biden's voice, suggesting New Hampshire voters should not vote (Associated Press 2024). A different US-related episode occurred during the 2016 US Presidential elections, when algorithms, automation, and GAI were used to increase the effectiveness and reach of disinformation campaigns and related cyber activities, which influenced the opinions and voting decisions of American citizens (Kertysova 2018). The combination of multiple modalities by generative models poses a challenge in the fight against false information, making it more difficult for both users and algorithms to distinguish between reality and deception (Xu *et al.* 2023). For this reason, countermeasures have been adopted at the user, platform and government-level. User-level resources include educational and verification websites, while social media platforms such as Facebook, Twitter, and YouTube actively label, delete, limit, pre-emptively counter, and provide information regarding misinformation (Waldemarsson 2020). The European Union's approach against disinformation seeks to find a balance between conflicting constitutional rights and freedoms that converge in the sphere of disinformation. This approach has led to the adoption of the Digital Services Act, which strictly

7 Definition: "Generative AI can be thought of as a machine-learning model that is trained to create new data, rather than making a prediction about a specific dataset. A generative AI system is one that learns to generate more objects that look like the data it was trained on". Retrieved from: https://news.mit.edu/2023/explained-generative-ai-1109.

8 Definition (Article 3, AI Act): "means AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful". Retrieved from: https://artificialintelligenceact.eu/article/3/.

regulates platforms involved in the dissemination of disinformation (Turillazzi *et al.* 2023). Additionally, the introduction of further regulatory measures aimed at collaborating with platforms to combat specific online content, such as the Enhanced Code of Practice on Disinformation, has played a significant role in shaping the European strategy against disinformation (Pamment 2020). This could be the first rigorous approach in tackling biases, and thus fake news, in defence of fundamental human rights against AI-driven threats. Bias is more related to discrimination, which violates human rights, as the following section will examine in greater depth.

Artificial intelligence and the right to non-discrimination

Generative AI, as previously mentioned, operates through algorithms, which are essential for streamlining numerous tasks that impact our daily lives, often on a scale beyond human capability. This is possible because algorithms work following computer instructions to convert inputs into outputs. Despite these 'superpowers' the potential for bias in algorithms is extremely high (Bozdag 2013). Even if machines and technologies appear impartial and neutral, they are not. In fact, people who develop algorithms may embody certain biases, which are easily transferable to machines. Therefore, the tendency of algorithms to generate discriminatory outcomes is not intrinsic to AI technology itself, but rather deeply rooted in societal factors influencing those who train these technologies (Nelson 2019).

Even if it seems that discrimination derives from a biased algorithm, it is necessary to distinguish between bias and discrimination. The term "bias" may have several meanings depending on the context. For the purposes of this article, it can be defined as: "unequal treatment based on protected characteristics, which includes discrimination and hate-motivated crimes" (Nelson 2019). Discrimination, on the other hand, refers to a situation where a person's disadvantage is based on one or more of these protected attributes (Nelson 2019).

Clearly, algorithmic bias results in discrimination when reliance on a protected characteristic causes a less favourable outcome, especially when coded parameters, training data, or input data feature elements that directly reflect a protected characteristic (European Union Agency for Fundamental Rights 2022). However, algorithmic bias most often leads to implicit discrimination, which is harder to identify and resolve.

Despite the evident risks that AI technologies pose, the use of algorithmic decision-making has been adopted in the U.S. criminal justice system to estimate potential recidivism. A prominent example of such a tool is the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) algorithm (Deb 2023). In 2016, the investigative journalism ProPublica conducted a thorough analysis of COMPAS, discovering racial biases in its risk assessments procedure. The original aim of COMPAS was to assist criminal justice professionals to assess possible future offenses (Brackey 2019). Even though COMPAS does not consider race as a factor to be calculated, in ProPublica's findings revealed disproportionate negative effects on Black defendants (Deb 2023). ProPublica examined over 7,000 cases between 2013 and 2014 to compare predicted recidivism scores with actual reoffending rates (Angwin *et al.* 2016). What emerged was deeply concerning: the algorithm's ability in predicting recidivism appeared to be very poor, only 20% of those labelled as high-risk actually reoffended (Angwin *et al.* 2016). The problem of discrimination emerged when the algorithm proved to be inaccurate for both black and white defendants. Black defendants were twice as likely to be wrongly classified as high-risk, whereas white defendants were more likely to be incorrectly classified as low-risk (Angwin *et al.* 2016).

Despite these alarming flaws, COMPAS was used in court sentencing decisions (Deb 2023). Afterwards, the founder of COMPAS acknowledged that the algorithm had not been designed for sentencing purposes, stating that: "the risk score alone should not determine the sentence of an offender" (Supreme Court 2016). Nonetheless, the indiscriminate use of this score in court trials raises profound ethical issues concerning automation bias, especially when human decision-makers rely exclusively on algorithmic results without critical evaluation.

To prevent such episodes, EU institutions have become increasingly engaged in addressing AI bias and other fundamental rights concerns. The AI Act (European Union 2024), a pioneering legislative proposal by the European Commission, regulates AI systems based on a risk-based framework. This

approach classifies AI applications into different categories, with the strictest regulations applied to high-risk systems, such as those used in employment, education, healthcare, and law enforcement. These systems must adhere to stringent requirements, including bias monitoring, transparency, and human oversight. The AI Act also prohibits certain AI practices outright, such as social scoring and manipulative techniques that could unfairly exploit individuals (*ibidem*, Article 5). Through these regulations, the EU aims to ensure that AI systems operate in an ethical, fair, and non-discriminatory manner. From another perspective, the General Data Protection Regulation (GDPR), which came into effect in 2018, plays a crucial role in shaping the ethical use of AI, particularly in the realm of automated decision-making (European Parliament and Council of the European Union 2016). Article 22 of the GDPR grants individuals the right to contest and opt out of fully automated decisions that have significant consequences on their lives, such as hiring processes or financial approvals (European Parliament and Council of the European Union 2016). Furthermore, the regulation enforces strict data protection principles, ensuring that personal data used in AI models is processed fairly and does not lead to discriminatory outcomes. Transparency is also a fundamental requirement under GDPR, as individuals have the right to be informed about how AI-driven decisions affecting them are made. The Digital Services Act (DSA)[9] and Digital Markets Act[10] (DMA) complement these efforts by targeting large digital platforms and ensuring that algorithmic decision-making processes remain transparent and accountable. These regulations aim to prevent discriminatory practices by compelling major technology companies to disclose how their AI systems operate, particularly in areas like content moderation, online advertising, and recommendation algorithms. By establishing these legal safeguards, the EU seeks to create a digital environment where AI-driven services respect fundamental rights and promote fairness.

Artificial intelligence and the right to privacy

In today's digital age the enormous quantities of data created and exchanged online allow businesses, governments, and organisations to improve their decision-making processes. Most of the time, this data includes sensitive information that individuals may prefer to keep private. This is where the concept of privacy comes into play. Privacy refers to the right to keep personal information secure and protected from unauthorised access (De Capitani Di Vimercati *et al.* 2012). This right has become increasingly significant due to the increasing volume of personal data being collected and analysed (DeVries 2003).

AI poses a risk to the privacy of both individuals and organizations, mainly due to the complexity of the algorithms used in AI systems and the fact that personal data is often a key component of the datasets used to train machine learning models. Additionally, AI technology is capable of making decisions based on subtle data patterns that often go unnoticed by humans. One of the most common issues consists in surveillance practices undermining personal autonomy and existing power disparities. This practice stems from the need to gather massive amounts of data to train AI systems, thus increasing the volume of personal data in AI systems and raising serious concerns about privacy and data protection.

Governments must therefore proactively safeguards individuals' privacy rights by implementing powerful data security measures, ensuring that data is used exclusively for its intended purposes. Furthermore, transparency in the use of personal data by AI systems is vital in order for users to understand how their data is being utilised.

Within the European Union framework, the adoption of the AI Act, specifically Article 78, imposes the confidentiality "of information and data obtained in carrying out their tasks and activities" (European Union 2024). The scope of this provision is to protect intellectual property rights, confidential business information, and trade secrets of both individuals and legal entities, including source code, as well as public and national security interests and the conduct of criminal or administrative proceedings. At another level, the General Data Protection Regulation (GDPR) outlines specific obligations for businesses

9 An official overview of the Act is provided by the European Commission at: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en.

10 An official overview of the Act is provided by the European Commission at: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en.

and organisations regarding the collection, storage, and management of personal data. It applies to both European organisations handling the personal data of individuals within the EU and to organisations based outside the EU that target individuals residing in the EU. Even though AI is not explicitly mentioned in the GDPR, AI systems must strictly comply with its requirements, especially when processing the personal data of EU citizens.

Numerous violations have occurred despite rigorous regulations designed to protect users. In 2022, Clearview AI, a facial recognition firm, became embroiled in one of the most significant privacy controversies in recent years. This company gathered billions of images from public websites without users consent to train its facial recognition technologies, enabling law enforcement agencies to match faces with identities into a database (Dul 2022). These practices raised serious concerns about violations of privacy rights, so much so that the Italian data protection authority launched an investigation following media reports highlighting multiple issues related to Clearview AÌs facial recognition products. In 2021, the Authority had also received further complaints from privacy and fundamental rights organisations expressing concern about Clearview AÌs activities. The investigative results highlighted several GDPR violations, including breaches of transparency, purpose limitation, and storage limitation principles (Pisanu 2022). Ultimately, the Italian data protection authority imposed a fine of EUR 20 million and the prohibition for any further web scraping activities within Italian territory. In addition, Clearview AI was required to appoint a representative within the European Union[11].

2. Introductory legal framework to Artificial intelligence

Until recently, authorities lacked the powers, procedural frameworks, and resources needed to ensure and monitor compliance with AI-related regulations, leaving national bodies solely responsible for enforcing safety and fundamental rights standards. This legal uncertainty and complexity discouraged businesses from developing and using AI technologies, slowing AI advancement in Europe and reducing the EUs global competitiveness in this sector. As already discussed, AI systems pose significant threats to fundamental rights such as non-discrimination, freedom of expression, human dignity, and the protection of personal data and privacy. The general framework of right protection under the EU Charter of Fundamental Rights and the General Data Protection Regulation (GDPR), was not sufficient to guarantee adequate safeguards against AI technologies (European Commission 2021). As a result, EU policymakers have committed to adopting a "human-centric" approach to AI, ensuring that Europeans can benefit from new technologies while upholding the EUs principles and values (Amariles and Baquero 2023). In its 2020 White Paper on Artificial intelligence (European Commission 2020), the European Commission pledged to promote the adoption of AI while also addressing the risks associated with certain applications of this emerging technology. Following an initial soft-law strategy, which included the 2019 Ethics Guidelines for Trustworthy AI and related policy and investment recommendations, the European Commission transitioned to a legislative approach (Madiega 2021). This development established standardised regulations for the development, market introduction, and application of AI systems (*ibidem*). The real breakthrough in AI legislation came in 2024 with the adoption of the AI Act, a targeted legislation capable of providing a clear legal framework to guide future AI development, ensuring compliance with ethical and legal standards while fostering innovation in a responsible manner.

Artificial intelligence act: the AI Act

In June 2024, European Union lawmakers signed the AI Act, establishing the first binding horizontal regulation on AI worldwide. The primary objective

11 A concise recap of the Clearview case is provided in a press release by the Italian Data Protection Authority, available at: https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9751323.
For the full decision issued by the Italian Authority against Clearview, see: Garante per la Protezione dei Dati Personali – GPDP (2022) *Ordinanza ingiunzione nei confronti di Clearview AI* - 10 febbraio [9751362], available at: https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9751362.
The Clearview case has also been referenced at the EU level in a news post by the European Data Protection Board (EDPB), available at: https://www.edpb.europa.eu/news/national-news/2022/facial-recognition-italian-sa-fines-clearview-ai-eur-20-million_en.

of this initiative is to ensure a correct functioning of the single market by fostering the development and deployment of trustworthy Artificial intelligence across the Union (European Union 2024).

The AI Act was designed as a comprehensive legislative framework that applies to all AI systems available or used within the EU. It is based on Article 114[12] and Article 16[13] of the Treaty on the Functioning of the European Union (TFEU). This framework aligns with the EUs strategy to ensure that products entering the EU market comply with existing legislation through conformity assessment and the CE marking system. However, due to its complexity, potential biases, and rapid evolution, AI often struggles to align with the legal requirements for certainty, transparency, accountability, and equal treatment. To address these challenges the EU legislator deliberately refrained from providing a rigid definition for "Artificial intelligence". Instead, the AI Act adopts a broad, nearly technology-neutral, and adaptable definition of "AI systems". In essence, the AI Act functions as a "Software Act". To reinforce this, the Commission incorporated a legal definition of "AI system" in Article 3 of the AI Act[14]. This definition encompasses various software-driven technologies that utilise specific methods and techniques, including "machine learning", "logic and knowledge-based" systems, and "statistical" methods. As stated in Article 2[15], the AI Act applies to providers who release or activate AI systems within the Union, regardless of whether they are based in the Union or outside of it; users of AI systems located within the Union and providers and users of AI systems located in a non-EU country when the AI system output is used within the Union (McCarthy 2007).

As a result, the AI Act does not extend to individual end-users, as the definition of a user in Art. 3(4)[16] explicitly excludes the use of AI systems in personal, non-professional contexts (Smuha *et al.* 2021). Accordingly, if the legislation can establish technical requirements for AI developers, it fails to establish robust and enforceable protection for fundamental rights. This happens due to weak provision, with several exceptions, aimed at safeguarding civic freedoms. Moreover, due to the need for flexibility in regulating evolving technologies, the final text of the Act contains several legal ambiguities and enforcement gaps, with many critical and specific aspects deferred to delegated acts, guidelines and future codes of conduct.

Nevertheless, the Commission has focused on adopting a risk-based regulation approach, grounded in the intended use of AI systems. Four categories of risk have been outlined[17]. At the top of this ranking system is the "unacceptable risk" category, which includes AI applications that undermine Union values and fundamental rights and are therefore prohibited. These prohibitions, listed in Article 5 (European Parliament and Council of the European Union 2024), include: subliminal techniques (resulting in physical and psychological harm); manipulative AI that leads to physical and psychological harm; social scoring and real-time remote biometric identification systems. The next category includes 'high-risk' AI technologies (Article 6), which are permitted but subject to compliance with specific AI requirements and ex-ante conformity assessments (Articles 16 and 43). Once these steps are successfully completed, the system can be affixed with the CE marking, signifying conformity with EU legislation. At this point, the AI application is ready to be introduced in the market or deployed for use. However, compliance does not end at market entry. A post-market monitoring procedure is also

12 Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12008E114.

13 "Article 16 of the Treaty on the Functioning of the European Union grants all individuals the right to the protection of their personal data. It also requires the European Parliament and the Council of the European Union to lay down rules to protect individuals with regard to the processing of personal data by EU institutions, bodies, offices and agencies, and by the Member States when carrying out activities that fall within the scope of EU law". Retrieved from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM:data_protection.

14 Definition: "Article 3 (1) of the AIA defines AI systems as software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with" (Ruschemeier 2023).

15 Available at: https://artificialintelligenceact.eu/article/2/.

16 Definition: "deployer means a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity". Retrieved from: https://artificialintelligenceact.eu/article/3/.

17 AI Act as risk-based approach regulation.

required (Article 72[18]). Providers must establish and maintain a post-market monitoring system tailored to the characteristics of the AI technologies, and to the risks associated with the high-risk AI system throughout their entire lifecycle.

In this phase, two key requirements come into play: data governance and human oversight. The first provision (Article 13) addresses concerns in relation to the quality of data employed in the development of AI systems, thus including rules on how training datasets should be structured and utilised along with guidelines for data preparation. Human oversight requirements (Article 14) ensure that AI systems are designed and developed in a way that enables effective human supervision throughout their operation period, focusing on "empowering human overseer" to detect irregularities, "automation bias," interpreting the system's outputs, and identifying faults where necessary.

The ultimate category addressed in the AI Act is the "limited risk" and the "minimal or no risk" AI systems. The former must be subject to information and transparency obligations, as outlined in Article 52[19]. This provision obliges providers to notify users when they are interacting with an AI system, if not clearly evident. Additionally, it is mandatory to apply labels to deep fakes so that users can easily recognise them.

"Minimal or no risk" AI systems are permitted without restrictions. However, the European Commission encourages the voluntary adoption of codes of conduct for AI with specific transparency requirements.

The EU AI Act presents both opportunities and challenges for European AI developers. On the one hand, meeting all the aforementioned requirements and administrative burdens could hinder the market entry of a product, consequently stifling innovation and research in this sector. On the other hand, the AI requirements are necessary to guarantee trust and safety, allowing users to interact with AI systems that meet high standards of transparency and security. Furthermore, the AI Act aligns with the GDPR, increasing and balancing the need for a high level of safety. Where the GDPR imposes stringent regulations on personal data protection, the AI Act governs the development and use of AI systems, including cases where personal data is not involved. Ultimately, the Act also extends to non-European companies operating in the EU market.

As Europe continues its digital transformation, both frameworks will play a crucial role in promoting innovation.

U.S. legal approach to Artificial intelligence

As AI-related legislation progresses around the world, the United States legislative activity highlights the nation's dedication to adapting to Artificial intelligence. The U.S. does not have a unified AI Act. Instead, its approach consists of a collection of policies designed to boost innovation while managing associated risks. The most recent one was released by US President Joe Biden, whose Executive Order on the "Safe, Secure, and Trustworthy Development and Use of AI" was adopted on 30 October 2023[20]. This Order aims to address algorithmic discrimination and securing voluntary commitments from leading US companies (such as Amazon, Google, Meta, Microsoft, and OpenAI) to foster safe, secure, and trustworthy AI advancement. The Order addresses different policy areas. Firstly, it emphasises the need for new standards regarding AI safety and security. For example, it mandates that developers of AI systems share their safety testing results and other important data with the US government; it safeguards citizens'privacy against AI-related threats, emphasising federal efforts to hasten the adoption and utilisation of privacy-preserving methods; it promotes equity and civil rights to defend consumers, patients, and students. This is just the latest Executive Order on the matter, although legislative proposals have been emerging since 2020, when the National Artificial intelligence Initiative Act provided guidance for AI research, development, and evaluation activities within federal science agencies[21]. Additional acts have mandated specific agencies to lead AI programs and policies throughout the federal

18 Available at: https://artificialintelligenceact.eu/article/72/.

19 Available at: https://artificialintelligenceact.eu/article/52/.

20 As also recalled in a one-year-later press release by the U.S. Department of Commerce (2024), *"Commerce Marks One-Year Anniversary of Historic Biden-Harris Initiative", a*vailable at: https://www.commerce.gov/news/press-releases/2024/10/commerce-marks-one-year-anniversary-historic-biden-harris.

21 U.S. Congress (2020), *AI in Government Act of 2020* (H.R. 2575), available at: https://www.congress.gov/bill/116th-congress/house-bill/2575.

government, such as the "AI in Government Act"[22] and the "Advancing American AI Act"[23].

Nonetheless, with the commencement of Donald Trump's second term, many of the initiatives introduced by Biden are now being revoked. In fact, President Trump enacted a new Executive Order called "Removing Barriers to American Leadership in Artificial intelligence"[24] This directive aims to eliminate rules viewed as hindering AI innovation, facilitating an "unbiased and agenda-free" approach to the development of AI technologies.

At the state level, jurisdictions such as Colorado, Illinois, and California are actively shaping the legislative environment regarding AI compliance for businesses. As an example, the state of Colorado, inspired by the EU AI Act, adopted a risk-based legislation to regulate the use of high-risk AI systems, emphasising transparency, risk management, and thorough documentation throughout AI system development (Siegal and Garcia 2024).

3. A Comparison of EU and U.S. Approaches to AI Regulation

As the AI sector grows, so do the risks associated with AI technologies, making it more challenging to draft legislation that clearly defines measures, obligations, and boundaries for businesses in this field (Software Improvement Group 2024). European legislators have opted for a human-centric approach recognising the multiple and severe risks that Artificial intelligence systems pose to human rights, safety, the rule of law, and democracy (*ibidem*). In contrast, the U.S. approach prioritises the private sector, avoiding stringent AI regulation and thereby allowing businesses to expand in the sector.

The present section provides a comparative analysis of the European and American approaches to AI legislation, exploring their differences, similarities, strengths, and weaknesses.

The AI Act and the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial intelligence present distinct approaches to the governance and regulation of AI. The AI Act is a regulation that applies directly and is legally binding in all EU Member States. In contrast, the Executive Order is a directive from the President of the United States that carries legal weight but influences policy and governance at the federal level without requesting Congressional approval.

When assessing the differences between the EU AI Act and the U.S. Executive Order, several immediate contrasts become apparent. The Executive Order primarily sets forth guidelines for federal agencies to follow, influencing industry practices. However, it does not impose regulations on private companies, except for specific reporting requirements under the Defense Production Act. On the other hand, the EU AI Act applies to any provider of general-purpose AI models operating within the EU, thereby having a much broader reach. While the U.S. Executive Order can be altered or revoked, especially with changes in presidential administrations, as seen with President Trump's election, the European AI Act establishes legally binding regulations within a more permanent governance framework. This difference highlights Europe's centrally coordinated and comprehensive regulatory framework, which can enact more binding rules compared to the U.S. approach. As noted, the U.S. Congress tends to leave more room for businesses to operate (Engler 2023). In contrast, the EU has placed a stronger emphasis on public transparency and on ensuring information about the role of AI in society. In this respect, the EU AI Act adopts a broader perspective on systemic risks, including issues such as widespread discrimination, significant accidents, and human rights implications. Both perspectives opted for a risk-based approach and similar principles in regulating AI, to ensure and develop a trustworthy AI system (*ibidem*).

Both legislative frameworks highlight the relevance of technical standards and best practices to achieve their goals. While the AI Act refers to them in Articles 17, 31, 32 and 40 – promoting a collaborative approach with European institutions and stakeholder organisations – the American Executive Order relies on federal agencies and expert bodies for implementation guidance, allowing the legal landscape to adapt over time. Therefore,

22 U.S. Congress (2020), *AI in Government Act of 2020* (H.R. 2575), available at: https://www.congress.gov/bill/116th-congress/house-bill/2575.

23 U.S. Congress (2021), *Advancing American AI Act* (S.1353), available at: https://www.congress.gov/bill/117th-congress/senate-bill/1353.

24 Available at: https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/.

this comparison is particularly relevant for global AI governance. By examining how two leading jurisdictions approach AI regulation, stakeholders can identify opportunities for dialogue, cooperation, and the development of shared norms. This process is vital not only to minimise regulatory conflicts but also to ensure that AI systems serve humanity in a safe, just, and inclusive manner.

Conclusion

Although Artificial intelligence seems to represent one of the greatest technological revolutions of our times due to its capability of transforming the way we interact with information, make decisions and manage critical processes, its implications extend far beyond scientific progress, raising fundamental questions related to human rights, transparency, and ethics. The issues analysed in this article – ranging from privacy concerns to discrimination and the spread of misinformation – demonstrate that the adoption of AI systems cannot occur without an adequate regulatory framework. The rapid development of AI has often outpaced lawmakers'ability to regulate its use, leading to scenarios where these technologies are deployed without sufficient safeguards for individuals. The case studies examined, including the risks associated with predictive algorithms in the judicial system and the proliferation of AI-generated content that distorts public discourse, highlight the urgent need for stronger regulatory intervention.

The adoption of the European Artificial intelligence Act has been a significant step in AI regulation, introducing a risk-based approach model capable of balancing technological innovation with the protection of fundamental rights, imposing stricter obligations on high-risk systems and banning unacceptable practices, such as social scoring and real-time biometric surveillance. The United States, on the other hand, has taken a more flexible and decentralised approach, with fragmented regulations across federal and state levels, influenced by the interests of the technology industry. The comparison between these two models highlights not only philosophical and legal differences but also the challenges of global AI governance.

Looking into the future, it seems to be necessary to develop an international regulatory ecosystem to coordinate the challenges posed by Artificial intelligence and ensure fairness, reliability, and security globally.

Moreover, scientific research must focus on ensuring AI systems' transparency and interpretability. In this respect, the notion of Trustworthy AI, promoted by the European Union, must translate into systems capable of explaining their decisions and operating responsibly. This implies the adoption of algorithmic auditing tools, increased accountability for AI providers, and stronger human oversight mechanisms (Laux *et al.* 2024).

This article has investigated the significant influence of Artificial intelligence on modern society. AI has progressed from its initial conception by McCarthy to the advanced neural networks used today. It represents a technological revolution due to its ability to analyse extensive datasets, identify intricate patterns, and automate decision-making processes that, in some tasks, surpass human performance. Nevertheless, this remarkable technological progress raises ethical and legal concerns regarding how AI systems function, how data is used, and the impact on individuals and communities. The analysis presented in the current article has highlighted the threats posed to essential human rights – particularly the right to privacy, access to reliable information, and protection against discrimination – providing clear examples of violations of these rights. The personalisation of algorithms driven by AI and deepfake technology contributes to the spread of misinformation, hindering individuals'access to accurate and diverse information. As previously stated, systems relying on automated decision-making can perpetuate biases, resulting in discriminatory practices in recruitment, law enforcement, and access to financial services. Additionally, AI surveillance tools, such as facial recognition, endanger privacy by facilitating extensive data gathering and intrusive monitoring. The risks highlighted emphasise the critical necessity for a regulatory framework that ensures AI development and implementation is with human rights standards.

The European Union is taking significant steps to mitigate the dangers arising from such technologies through initiatives like the AI Act, which specifically regulates high-risk AI applications while promoting transparency, accountability, and fairness within AI systems. Nonetheless, obstacles remain in finding a balance between innovation, ethical considerations

and legal obligations. Ultimately, in order to fully harness AÌs potential, it must be developed in a way that prioritises trust and protects fundamental rights. This raises the necessity of ensuring the robustness of AI systems, so that they continue to function as intended even when their inputs or models are subjected to adversarial interference. Assessing the robustness of a system requires testing for any possible input perturbations. However, this is practically impossible for AI, as the number of possible interferences is often exorbitantly large. If AI robustness cannot be adequately assessed, neither can its trustworthiness (Tocchetti *et al.* 2022).

Philosophical analyses define trust as the decision to delegate a task without any form of control or supervision over the way the task is executed. Successful instances of trust rest on an appropriate assessment of the trustworthiness of the agent to which the task is delegated (the trustee). Therefore, trustworthiness can be considered both a prediction of the likelihood that the trustee will behave as expected, based on past behaviour, and a measure of the risk borne by the trustor should the trustee fail to do so (Li *et al*. 2023). According to the International Organisation for Standardisation, trustworthy AI is "a framework to ensure that a system is worthy of being trusted based on the evidence concerning its stated requirements. It ensures that users'and stakeholders'expectations are met in a verifiable way (ISO/IEC 2020)".

The lack of transparency and the learning abilities of AI systems, along with the nature of attacks on these systems, make it difficult to evaluate whether the system will continue to behave as expected in any given context. Thus, it is important to de-anthropomorphise AI; AI systems cannot be held as responsible agents because no penalty imposed on them would motivate them to improve their behaviour (Li and Suh 2021). Policies should mandate a cautious evaluation of the level of control over AI systems based on the tasks they are designed to perform. In conclusion, the concept of reliability should be preferred over trustworthiness. A reliable AI system would require constant human oversight, ensuring that human operators understand how to approach different levels of automation. In particular, it is necessary to guard against the "half-automation problem": the phenomenon in which, when tasks are highly but not fully automated, the human operator tends to rely on the AI system as if it were fully automated. This underscores the importance of incorporating "kill switches" to maintain human control over AI processes, especially in the field of automated responses (Polito and Pupillo 2024).

The European Union has moved in this direction with the 2019 publication of the "Ethics Guidelines for Trustworthy Artificial intelligence," which established seven key requirements that AI systems should meet to be deemed trustworthy: human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination, fairness, societal and environmental well-being, and accountability. These guidelines urge all parties involved to adopt these seven essential criteria and to promote the Union's focus on human-centric principles (European Commission 2019). Nevertheless, to date, it remains vital to maintain reliable AI systems, thus ensuring human supervision to reduce risks and preserving authority over automated operations (Floridi 2021). An approach centred on human values, as encouraged by the EU, is key to aligning AI advancement with ethical and legal guidelines, cultivating trust and responsibility.

References

Amariles D.R., Baquero P.M. (2023), Promises and limits of law for a human-centric artificial intelligence. *Computer Law & Security Review*, *48*, 105795 <https://doi.org/10.1016/j.clsr.2023.105795>

Angwin J., Larson J., Mattu S., Kirchner L. (2016), *How we analyzed the COMPAS recidivism algorithm,* ProPublica <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>

Associated Press (2024), Company that sent fake Biden robocalls in New Hampshire agrees to $1m fine, *The Guardian*, 22 August

Blank I.A. (2023), What are large language models supposed to model?, *Trends in Cognitive Sciences*, 27, n.11, pp.987-989

Block H.D. (1962), The perceptron: A model for brain functioning, *Reviews of Modern Physics*, 34, n.1, 123-135

Bontridder N., Poullet Y. (2021), The role of artificial intelligence in disinformation, *Data&Policy*, 3, n.e32 <doi:10.1017/dap.2021.20>

Bozdag E. (2013), Bias in algorithmic filtering and personalization, *Ethics and information technology*, n.15, pp.209-227

Brackey A. (2019), *Analysis of racial bias in Northpointe's COMPAS algorithm*, Master's thesis at Tulane University https://library.search.tulane.edu/discovery/fulldisplay/alma9945513177106326/01TUL_INST:Tulane

Couldry N., Mejias U.A. (2019), Data colonialism: Rethinking big data's relation to the contemporary subject, *Television & New Media*, 20, n.4, pp.336-349

Council of Europe (2024), *Council of Europe Framework Convention on Artificial intelligence and Human Rights, Democracy and the Rule of Law*, Council of Europe Treaty Series n.225, Vilnius, Council of Europe

De Capitani Di Vimercati S., Foresti S., Livraga G., Samarati P. (2012), Data privacy: Definitions and techniques, *International Journal of Uncertainty*, *Fuzziness and Knowledge-Based Systems*, 20, n.06, pp.793-817

Deb E. (2023), COMPAS — an AI tool sending or keeping people in Jail, *edblogs.medium.com*, december 24

DeVries W.T. (2003), Protecting privacy in the digital age, *Berkeley Technology Law Journal*, 18, n.1, pp.283-311

Dul C. (2022), Facial recognition technology vs privacy: The case of Clearview AI, *The Queen Mary Law Journal*, 23, pp.1-24

Engler A. (2023), *The EU and U.S. diverge on AI regulation: A transatlantic comparison and steps to alignment, Brookings Institution*, United States of America, Retrieved from https://coilink.org/20.500.12592/3ptdkj on 21 Jul 2025

European Commission (2021), *Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial intelligence Act) and amending certain Union legislative acts* COM/2021/206 final

European Commission (2020), *White paper on artificial intelligence: A European approach to excellence and trust* COM(2020) 65 final

European Commission (2019), *Building trust in human-centric artificial intelligence,* COM(2019) 168 final

European Parliament, Council of the European Union (2024), *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending various regulations and directives (Artificial intelligence Act) (PE/24/2024/REV/1),* Official Journal of the European Union

European Parliament, Council of the European Union (2016), *Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Official Journal of the European Union,* L 119, 1-88 http://data.europa.eu/eli/reg/2016/679/oj

European Union (2024), *Artificial intelligence Act*, Official Journal of the European Union, L 1689, 12 July https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

European Union Agency for Fundamental Rights (2022), *Bias in algorithms – Artificial intelligence and discrimination*, Luxembourg, Publications Office of the European Union

Festré A., Garrouste P. (2015), The 'Economics of Attention': A History of Economic Thought Perspective, *OEconomia,* 5, n.1, pp.3-36

Floridi L. (2021), The european legislation on ai: A brief analysis of its philosophical approach, *Philosophy & Technology*, 34, n.2, pp.215-222

Hardesty L. (2017), Explained: Neural networks. Ballyhooed artificial-intelligence technique known as "deep learning" revives 70-year-old idea, *MIT News,* April 14

Hsu T. (2024), New Hampshire Officials to Investigate A.I. Robocalls Mimicking Biden*, The New York Times*, January 22

ISO/IEC (2020*), TR24028:2020 Information Technology – Artificial intelligence – Overview of Trustworthiness in Artificial intelligence [Standard],* Vernier (Geneva), International Organization for Standardization

Kenaw S. (2008), Hubert l. Dreyfus's critique of classical ai and its rationalist assumptions, *Minds and Machines*, 18, n.2, pp.227-238

Kertysova K. (2018), Artificial intelligence and Disinformation: How AI Changes the Way Disinformation is Produced, Disseminated, and Can Be Countered, *Security and Human Rights,* 29, n.1-4, p.55-81

Laux J., Wachter S., Mittelstadt B. (2024), Trustworthy artificial intelligence and the European Union AI Act: On the conflation of trustworthiness and acceptability of risk, *Regulation & Governance*, 18, n.1, pp.3-32

LeCun Y., Bengio Y., Hinton G. (2015), Deep learning, *Nature*, 521, n.7553, pp.436-444 https://doi.org/10.1038/nature14539

Li B., Qi P., Liu B., Di S., Liu J., Pei J., Yi J., Zhou B. (2023), Trustworthy ai: From principles to practices, *ACM Computing Surveys*, 55, n.9, pp.1-46

Li M., Suh A. (2021), Machinelike or humanlike? A literature review of anthropomorphism in AI-enabled technology, in HICCS, *Proceedings of the 54th Hawaii international conference on system sciences*, Kauai (Hawaii, USA), ScholarSpace, pp.4053-4062

Madiega T. (2021), *Artificial intelligence act,* Bruxelles, European Parliamentary Research Service

Mantelero A. (2022), *Beyond data: Human rights, ethical and social impact assessment in AI*, London, Springer Nature

McCarthy J. (2022), Artificial intelligence, logic, and formalising common sense, in Carta S. (ed.), *Machine Learning and the City: Applications in Architecture and Urban Design*, Bognor Regis (UK), John Wiley & Sons Ltd., pp.69-90

McCarthy J. (2007), *What is Artificial intelligence*, Stanford (CA), Stanford University

Neapolitan R.E. (2012), *Contemporary Artificial intelligence*, Boca Raton (USA), Chapman and Hall/CRC

Nelson G.S. (2019), *Bias in artificial intelligence*, *North Carolina medical journal*, 80, n.4, pp.220-222

Pamment J. (2020), *The EU Code of Practice on Disinformation: Briefing Note for the New EU Commission*, Policy perspective series n.1, Washington (DC), Carnegie Endowment for International Peace

Pisanu N. (2022), Il caso Clearview AI, il riconoscimento facciale viola la nostra privacy: la multa del Garante, *Cybersecurity 360*, 9 marzo

Polito C., Pupillo L. (2024), Artificial intelligence and cybersecurity, *Intereconomics,* 59, n.1, pp.10-13

Rombach R., Blattmann A., Lorenz D., Esser P., Ommer B. (2022), High-resolution image synthesis with latent diffusion models, in *CVPR 2022. 2022 Conference on Computer Vision and Pattern Recognition (CVPR)*, Danvers (MA), Clearance Center, pp.10674-10685

Rosenblatt F. (1958), The perceptron: A probabilistic model for information storage and organization in the brain, *Psychological Review*, 65, n.6, pp.386-408

Ruschemeier H. (2023), AI as a challenge for legal regulation–the scope of application of the artificial intelligence act proposal, *Era Forum*, 23, n.3, pp. 361-376

Seow J.W., Lim M.K., Phan R. C.W., Liu J.K. (2022), A comprehensive overview of Deepfake: Generation, detection, datasets, and opportunities, *Neurocomputing*, 53, pp.351-371

Siegal A., Garcia I. (2024), A deep dive into Colorado's artificial intelligence act, *Attorney general Journal*, October 26

Sohail S., Farhat F., Himeur Y., Nadeem M., Madsen D.Ø., Singh Y., Atalla S., Mansoor W. (2023), *The Future of GPT: A Taxonomy of Existing ChatGPT Research, Current Challenges, and Possible Future Directions*, 8 April <http://dx.doi.org/10.2139/ssrn.4413921>

Smolensky P. (1987), Connectionist AI, symbolic AI, and the brain, *Artificial intelligence Review*, 1, n.2, pp.95-109

Smuha N., Ahmed-Rengersb E., Harkens A., Li W., MacLaren J., Piselli R., Yeung K. (2021), *How the EU can achieve legally trustworthy AI: a response to the European Commission's Proposal for an Artificial intelligence Act*, Birmingam, LEADS Lab @ University of Birmingam <http://dx.doi.org/10.2139/ssrn.3899991>

Software Improvement Group (2024), *AI legislation in the US: A 2025 overview* https://www.softwareimprovementgroup.com/us-ai-legislation-overview/

Tocchetti A., Corti L., Balayn A., Yurrita M., Lippmann P., Brambilla M., Yang J. (2022), Ai robustness: a human-centered perspective on technological challenges and opportunities, *ACM Computing Surveys*, 57, n.6, Article 141

Treusch P. (2020), Re-reading ELIZA: Human–machine Interaction as cognitive sense-ability, in Lorenz-Meyer D., Treusch P., Liu X. (eds.), Feminist Technoecologies. *Reimagining Matters of Care and Sustainability,* London, Routledge, pp.61-76

Turillazzi A., Taddeo M., Floridi L., Casolari F. (2023), The digital services act: an analysis of its ethical, legal, and social implications, *Law, Innovation and Technology*, 15, n.1, pp. 83-106

Turing A.M. (1950), Computing Machinery and Intelligence, *Mind*, n.49, pp.433-460

Waldemarsson C. (2020), *Disinformation, Deepfakes & Democracy: The European response to election interference in the digital age*, Copenhagen, The Alliance of Democracies Foundation

West D.M. (2018), What is artificial intelligence?, *Brookings.edu*, October 4

Xu D., Fan S., Kankanhalli M. (2023), Combating misinformation in the era of generative ai models, in *Proceedings of the 31st ACM International Conference on Multimedia*, New York, Association for Computing Machinery, pp.9291-9298

Laura Evangelista
l.evangelista@inapp.gov.it
Researcher, Head of the Research Group Accreditation and Quality for VET at INAPP. National Coordinator of the EQAVET National Reference Point for Italy. She coordinates research projects related to the monitoring and evaluation of ccreditation systems of VET providers in Italy, the implementation of quality arrangements in line with EU-evel principles and indicators, the experimentation of Peer Review at provider and system level, and the dissemination of results through seminars, conferences and papers. Recent publications: (with Fonzo C.), Self-assessment in VET and Higher education: links and further developments, *Quaderni di Comunità*, n.3, 2023; (with Fonzo C.), La metodologia europea della Peer Review: prima sperimentazione tra istituti scolastici e Centri di Formazione Professionale, *Rassegna CNOS*, n.39, 2023.

Concetta Fonzo
c.fonzo@inapp.gov.it
Deputy Coordinator of the EQAVET National Reference Point for Italy and Deputy Coordinator of ReferNet Italy. Member of INAPP's research group Accreditation and Quality for Training and of its National Operational Programmes funded by the ESF+. Since 2004, she has been a member of scientific and technical committees and boards for various European initiatives and networks. An expert in European project management, she contributes to national and international research activities in the fields of education, training, guidance and social inclusion. Recent publications: (with Evangelista L.), Self-assessment in VET and Higher education: links and further developments, Quaderni di Comunità, n.3, 2023; (with Evangelista L.), La metodologia europea della Peer Review: prima sperimentazione tra istituti scolastici e Centri di Formazione Professionale, *Rassegna CNOS*, n.39, 2023.

Eleonora Zecca
ele.zecca00@gmail.com
She holds a Master's degree in Policies and Governance in Europe from LUISS Guido Carli and a Bachelor's degree in Political Science and International Relations from the University of Parma. She also studied at the University of Passau (Germany) as part of a double degree experience. Her research interests focus on labor market dynamics and political systems. She authored her Master's thesis on *The Italian Labor Market and the Green Transition in SMEs. A Focus on the Automotive Sector* and her Bachelor's thesis on *Polarizzazione affettiva e sistemi di partito. Un'analisi comparata*.

IA: lo human first e il processo amministrativo

Andrei Mihai Pop
Università degli Studi di Milano - Dottorando

Enrico Ferraris
PagoPA SpA

La rivoluzione dell'Intelligenza artificiale rappresenta uno spartiacque per la storia dell'Uomo. In particolare, uno dei settori che subirà influenze sarà quello della Pubblica amministrazione. Dopo aver preliminarmente definito il concetto di IA, si introduce il 'nuovo' diritto allo *human first*, che con la diffusione delle soluzioni basate su IA dovrà essere sempre più preminente. Infine, si analizza l'impatto che l'IA può avere sulla PA, in particolare nell'ambito del procedimento amministrativo, e come questa dovrà reagire per continuare ad operare nel rispetto dei diritti dei cittadini.

The Artificial intelligence revolution marks a watershed moment in human history. In particular, one of the sectors that will be significantly impacted is Public administration. After initially defining the concept of AI, the article introduces the 'new' right to what is known as 'human first', which, with the spread of AI-based solutions, will need to become increasingly prominent. Finally, the analysis focuses on the impact that AI may have on public administration, particularly in the context of administrative procedures, and how the latter must respond in order to continue operating in compliance with citizens' rights.

DOI: 10.53223/Sinappsi_2025-02-7

Citazione
Pop A.M., Ferraris E. (2025), IA: lo human first e il processo amministrativo, *Sinappsi*, XV, n.2, pp.90-100

Parole chiave
Diritti civili
Intelligenza artificiale
Pubblica amministrazione

Keywords
Civil rights
Artificial intelligence
Public administration

1. Una definizione europea di IA

Sul piano europeo, il principale contributo relativo alla definizione dell'IA è ascrivibile alle iniziative dei principali organi dell'Unione europea, la Commissione e il Parlamento. È la prima, infatti, che diede, inizialmente, l'impulso per la ricerca di una condivisa ed efficace definizione di IA. Il primo risultato di tali sforzi è contenuto in una Comunicazione intitolata *Artificial Intelligence for Europe* del 25 aprile 2018 nel quale le istituzioni europee manifestarono i propri obiettivi strategici in materia di IA, concentrandosi sulle opportunità e le sfide poste da questa tecnologia. In sintesi, questi possono essere suddivisi in due macro-tematiche: da un lato, il tentativo di divenire una potenza globale, incoraggiando gli enti pubblici e i soggetti privati a investire sull'IA, dall'altro la necessità di farlo nel quadro di un sistema legale ed etico che ponesse al centro del discorso una IA *trustworthy* e *human-centered*[1]. Tanto premesso, è opportuno sottolineare come questo primo approccio non rimase a lungo

Il presente lavoro è frutto di riflessioni condivise degli Autori, tuttavia il paragrafo 2 e 3 sono da ricondurre esclusivamente e unicamente ad Andrei Mihai Pop, mentre il paragrafo 1 ad Enrico Ferraris.

1 Si veda il paragrafo III di detto comunicato.

'operativo'. Nello specifico, infatti, può dirsi superato poco più di un anno dopo dai lavori dell'AI HLEG (un gruppo di esperti istituito dalla Commissione dell'UE nel 2018), il quale elaborò una nuova definizione di IA ad aprile del 2019[2]. Pur potendo considerare la fonte autorevole, questa definizione non ebbe effetti cogenti e vincolanti, non essendo considerabile come un atto legislativo. Invero, lo scopo degli esperti non era quello di regolare la materia, bensì di aprire un dibattito su tale nuova tecnologia. Il punto di svolta è rappresentato dalla proposta da parte della Commissione di un Regolamento, ossia l'AI Act, dell'aprile 2021[3], proposta che ha evidentemente tratto giovamento dal contributo del c.d. AI HLEG[4]. Il Regolamento, a seguito di una lunga gestazione incominciata fin dalla prima Comunicazione da parte della Commissione UE, venne approvato il 21/5/2024 e successivamente pubblicato il 12/7/2024 sulla Gazzetta ufficiale dell'UE (Regolamento UE, n. 1689/2024). In tale testo legislativo si elabora, a livello europeo, la prima definizione ufficiale e vincolante di IA. Infatti, l'articolo 3 comma 1 dell'AI Act statuisce che:

> *AI system means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments*[5].

Ai fini di una completa comprensione di tale definizione, è imprescindibile analizzare anche il *Considerando* n. 12 del Regolamento. In tale disposizione, il Legislatore dell'UE ha inteso chiarire alcuni aspetti che potrebbero rimanere 'incompresi' leggendo isolatamente l'articolo 3, comma 1, stante l'opacità connaturata nel linguaggio umano e la difficoltà definitoria della materia.

In primo luogo, in tale *Considerando*, si precisa che il Regolamento non si applica ai sistemi fondati unicamente su regole poste da persone fisiche che comportino l'esecuzione di operazioni in automatico. Questa è, a tutta evidenza, una formula che mira ad escludere dall'applicazione dell'AI Act alcune categorie di prodotti che non possono dirsi basati su Intelligenza artificiale. In tale prospettiva, il citato *Considerando* prevede che il Regolamento debba applicarsi esclusivamente in presenza di determinate caratteristiche tecniche. Invero, i sistemi di IA si fondano su almeno cinque caratteristiche che li differenziano dai tradizionali software: 1) capacità inferenziale; 2) approcci di apprendimento automatico; 3) approcci di logica; 4) autonomia; 5) adattabilità.

Ovviamente, la corretta e concreta declinazione di tali principi e regole dipenderà dalle interpretazioni che gli operatori del diritto ne daranno. Dunque, si può affermare che l'UE abbia deciso di favorire una omogenea prospettiva atlantista con la finalità di coordinarsi anche con gli orientamenti dei principali partner, specialmente gli Stati Uniti d'America. È evidente che riuscire a porsi tra le locomotive nella corsa allo sviluppo dell'IA dovrebbe poter garantire immensi vantaggi economici[6].

2 Si veda *Ethics Guidelines for Trustworthy Artificial Intelligence* (European Commission 2019).

3 La proposta è disponibile al seguente link: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206 (consultato il 2/3/2025).

4 Purtuttavia, la definizione dell'AI Act non riproduce meramente le riflessioni dell'AI HLEG, bensì la reinterpreta, sviluppando e chiarendo alcuni aspetti. A titolo esemplificativo, si abbandona il riferimento tecnico al software o all'hardware preferendo la nozione di *machine-based system*.

5 L'Italia ha positivizzato – con la legge n. 132/2025 – una definizione, tuttavia è una mera riproposizione dell'indirizzo dell'UE non potendosi osservare alcuna differenza. Così anche Cassano (2024, 388).

6 L'approccio dell'AI Act mira ad avere gli stessi effetti che ebbe il GDPR. In tema di IA, si veda Finocchiaro (2024a, 113-125). Questo impatto dell'UE – prima col GDPR e ad oggi posto come obiettivo con l'AI Act – sugli altri attori internazionali è sintetizzabile con il c.d. *Brussels Effect*. Si veda Bradford (2020). Tuttavia, vi sono opinioni che ritengono più probabile un c.d. *Reverse 'effetto Bruxelles'*, ovverosia un forte rallentamento della crescita tecnologica nell'UE a causa dell'eccessiva regolamentazione. Inoltre, si sostiene che in realtà, almeno per quanto riguarda il 'mondo digitale' (contrapposto a quello 'fisico'), vi sarebbe non un 'effetto Bruxelles', bensì un 'effetto fotocopia'. In altre parole, l'Europa, in tale prospettiva, non riuscirebbe ad esportare la propria potenza normativa, ma assumerebbe solamente il ruolo di 'apripista' consentendo agli altri attori di trarre spunto, ferma però la possibilità di discostarsene. Si veda l'intervista del 26/8/2024 di Luciano Floridi disponibile al link: https://www.fondazioneleonardo.com/videos/effetto-bruxelles-normativa-europea-influenza-mondo-luciano-floridi (consultato il 4/3/2025); similmente Finocchiaro (2024b, 21-22). Per una prima analisi dei caratteri essenziali dell'AI Act – in particolare della governance – si rimanda anche a Pop (2024).

In definitiva, con questo AI Act, il vecchio continente mira a tracciare una 'terza via' che contemperi le esigenze economiche di sviluppo di una nuova tecnologia con le istanze etiche e giuridiche[7]. Una 'via europea' che si vorrebbe porre in alternativa rispetto all'impostazione legata al liberismo degli USA e a quella tipica dell'autoritarismo della Cina[8].

2. Lo human first come diritto fondamentale

Principiamo osservando quanto sia fondamentale porre al centro delle riflessioni, in tema dell'IA, l'essere umano. L'Uomo, con il suo ingegno, si è prodigato per migliorare le proprie condizioni di vita con l'uso-potenziamento della tecnica. Inizialmente, quest'ultima rimaneva strutturalmente uno strumento e mezzo asservito agli scopi del proprio creatore. Tuttavia, ad oggi, lo sviluppo della tecnica impone un sempre maggiore grado di autonomia delle invenzioni per incrementare il grado di soddisfazione della *societas*. Ciononostante, è doveroso tentare un contenimento di tale slittamento esistenziale attraverso lo strumento del diritto, il quale rimane una delle opere più longeve, potenti e 'umane'. La tecnica è presente nella vita dell'Uomo fin dai suoi albori: già l'uso di un bastone per procacciare cibo, o per offendere altri esseri, rappresentava la nascita di tale 'forza'. Purtuttavia, la *téchne* ha un gemello, nato in un periodo contiguo e, pressoché, in contemporanea. Ci si riferisce allo ius, il quale è già rinvenibile nelle prime regole di condotta di un dato gruppo di uomini rispetto ai propri membri (su un piano interno) e nei confronti di altri gruppi (su un piano esterno).

Queste due sono le 'forze' che indubbiamente sono presenti nella vita dell'Uomo fin dal principio e probabilmente lo accompagneranno per sempre. La tecnica come massima espansione delle ambizioni e potenze, alle volte incontrollate, dell'umanità; al contrario, lo ius come limite e guida verso un determinato scopo-fine che tenti di tutelare i diritti[9].

Negli ultimi decenni tale rapporto non è mutato, tuttavia la tecnica ha subìto una netta accelerazione e il diritto non è riuscito sempre ad adattarsi, anche a causa degli interessi economici che ne derivano e che impongono uno sviluppo sempre più rapido. Invero, sono addirittura sopraggiunte tesi che vorrebbero la sostituzione di uno ius 'umano' con un c.d. ius 'sintetico', in quanto oggettivo, imparziale e non più succube delle emozioni tipiche degli esseri umani. In tale ultimo scenario, la stessa esistenza del diritto, così come lo si conosce, rischierebbe di essere elisa[10]. Dunque, già da queste considerazioni iniziali, si può cogliere la necessità di respingere la tesi che vorrebbe un diritto assolutamente 'neutrale'[11] allo scopo di favorire il mercato, poiché una delle sue essenze è proprio quella di controbilanciare e di consentire la salvaguardia dei diritti degli individui. In tale prospettiva, quello che potrebbe essere ammesso è l'enucleazione di un insieme di norme coordinate per evitare 'fughe normative', ma non la sottomissione integrale dello ius alle esigenze mercantili. Questo parrebbe essere stato l'approccio dell'UE con l'AI Act emanato con l'intento di creare un mercato unico europeo anche nel campo dell'Intelligenza artificiale[12].

In tale contesto problematico si colloca l'emersione dell'IA, delle sue potenzialità e dei suoi rischi per i diritti fondamentali dell'Uomo. Questa tensione tra il dominio umano sulle proprie opere

7 Va però sottolineato che questa 'terza via' è una regolatoria e non produttivo-tecnologica. In altri termini, l'UE non ha la struttura e potenza di produzione necessaria per competere con gli Stati Uniti o con la Cina. Per tale ragione, l'UE sta tentando di spostare il 'campo di battaglia' sul piano normativo. Così Finocchiaro (2024a, 109).

8 Si veda Finocchiaro (2024a, 107-108). Questa impostazione mira a competere indubbiamente con l'America, ma specialmente con la Cina al fine evitare che il c.d. *Brussels Effect* venga sostituito da quello che è stato definitivo come il *Beijing effect*. Si rimanda per il *Beijing effect* a Erie e Streinz (2021).

9 Questa considerazione impone, altresì, che il diritto intervenga 'a monte' senza dover inseguire lo sviluppo costante e imprevedibile dell'Intelligenza artificiale. In altri termini, la portata e la forza del diritto deve inserirsi in un approccio by design all'IA che riesca a 'giocare d'anticipo' evitando, laddove possibile, di giungere al tradizionale approdo violazione-sanzione. Si veda Casonato (2019, 725).

10 Questa prospettiva pare non tenere sufficientemente in considerazione la circostanza per la quale, allo stato attuale, l'IA non sarebbe altro che un nostro riflesso, quindi con qualità, ma anche con difetti. Dunque, in tal guisa, la 'neutralità' dell'AI non sarebbe che un miraggio. Si veda Pietropaoli (2020, 110-113).

11 Sostiene tale tesi Abbott (2020, 134-143). Al contrario, Cal (2023, 65) vede nell'intervento del Legislatore europeo con l'AI Act un approccio "sostanzialmente neutro".

12 Il *Considerando* numero 1 dell'AI Act afferma che: "*The purpose of this Regulation is to improve the functioning of the internal market by laying down a uniform legal framework in particular for the development, the placing on the market, the putting into service and the use of artificial intelligence systems (AI systems) in the Union*".

e la ricerca continua di 'miglioramenti' della propria esistenza impone un contemperamento serio e stabile che solamente il diritto può svolgere.

In tale elaborato si accoglierà la tesi che impone la preferenza di un Uomo-dominus, rifiutando la deriva totale che impone di mutare la visione di un'umanità al centro del mondo, passando da una prospettiva tendenzialmente umano-centrica a una tecno-centrica. In tal guisa, la visione umano-centrica è stata cristallizzata anche dal Regolamento sull'Intelligenza artificiale che ha voluto imporre un'accelerazione su tali temi indicando una netta preferenza. Solamente in tal senso si può giustificare razionalmente l'enunciazione del diritto a una *AI Disclosure*, *Explainability* e *Transparency*. L'essere umano, dunque, deve mantenere un certo grado di controllo sull'IA e sul suo funzionamento al fine di tutelare i propri stessi diritti[13]. Questo criterio di gestione ha differenti gradi di 'potenza': la pervasività del controllo dell'Uomo rispetto alle operazioni compiute dall'IA aumenta all'aumentare dei diritti in gioco. Tali livelli di intervento sono conosciuti con le seguenti espressioni: *Human in command* (HIC), *Human in the loop* (HITL), *Human on the loop* (HOTL) e *Human out of the loop* (HOUTL)[14]. L'insieme di tali espressioni contribuisce alla formazione ed estensione del principio dello *human oversight*, ossia il controllo umano sul funzionamento dell'IA. Incominciando da quello più 'debole', ossia l'HOUTL, si può facilmente comprendere come non può trovare applicazione nel caso in cui siano in gioco i diritti delle persone: in questo scenario, infatti, vi è un'assenza totale dell'intervento umano, ovverosia il dominus assoluto è l'IA. Un simile livello di pervasività dell'IA appare, in via astratta, incompatibile con la maggior parte dei trattamenti che riguardino – direttamente o indirettamente – prerogative degli individui, salvo che i rischi siano nulli o minimi. Procedendo nell'analisi, e salendo nella 'scala dell'intervento umano', si giunge allo HOTL. In tale caso, l'IA ha l'iniziativa e la gestione del compito, mentre l'Uomo assume il ruolo di sorvegliare i suoi procedimenti potendo intervenire in una fase ex post rispetto all'output del sistema. Questo approccio può essere adottato laddove il sistema di IA non comporti pericoli rilevanti (rischi medio-bassi) per i diritti fondamentali. Invece, un grado di *human oversight* che consenta un più penetrante controllo-intervento dell'essere umano è quello dell'HITL che implica un controllo dell'essere umano dell'attività da svolgere: l'IA assiste, non decide. Questo scenario pare compatibile con la gestione della maggior parte dei casi d'uso che riguardino diritti fondamentali e che comportino un rischio medio-alto.

In ultimo luogo, si trova il c.d. HIC. In tale ipotesi, l'IA avrà un ristrettissimo ambito applicativo essendo l'intero procedimento controllato e guidato dall'Uomo. Contrariamente all'HOUTL, in tale caso è proprio l'Uomo ad essere il dominus. Tale rigore trova applicazione laddove le attività da svolgere possono coinvolgere e influenzare diritti fondamentali con un rischio alto-inaccettabile[15] – ad esempio, l'ambito della Giustizia. Dunque, il diritto allo *human oversight* deve essere declinato in base allo stato tecnologico e all'interesse oggetto di bilanciamento. Da un lato, più un sistema di IA è trasparente e spiegabile e maggiori sono le possibilità per seguire un approccio HOTL; dall'altro lato, assumendo una differente prospettiva, all'aumento di criticità del diritto oggetto di 'attenzione', vi è la necessità di utilizzare sistemi trasparenti e spiegabili. A prescindere dall''etichetta', tali espressioni possono essere ricondotte al diritto-dovere di applicare una sorveglianza umana sull'IA. Invero, questo criterio impone la regola secondo la quale in caso di utilizzo di tali sistemi, in settori giudicati come 'critici', occorrerà la supervisione dell'Uomo. In altri termini, l'Uomo e l'IA dovranno interagire e collaborare per giungere ad un output – c.d. *collaborative intelligence*[16]. Questo criterio-guida ha avuto già nel Regolamento europeo sulla protezione dei dati personali una sua iniziale, benché parziale, enunciazione all'articolo 22 comma 1, in materia di decisioni basate su trattamenti automatizzati. Ad una prima lettura, tale previsione risulta essere chiara e perentoria pur potendosi applicare solamente nei casi in cui l'IA utilizzi dati che possano essere qualificati come 'personali'. Inoltre, anche la prima parte del

13 In tali termini, pur riferendosi al c.d. elaboratore elettronico (potenzialmente dotato di IA), si veda Frosini (1983).

14 Propongono questa categorizzazione le *Ethics Guidelines for Trustworthy Artificial Intelligence* (European Commission 2019); per una prima interpretazione, si consenta di rimandare a Laviola (2020).

15 Abbiamo associato le varie categorie dell'*human oversight* con il rispettivo grado di rischio in base anche al *risk-based approach* dell'AI Act: HOUTL-nullo; HOTL-basso/medio; HITL-medio/alto; HIC-alto/inaccettabile.

16 In termini simili, Quintarelli *et al.* (2019, 191 e 199).

Considerando n. 71 del GDPR pare rinforzare tale prospettiva. Purtuttavia, tali enunciazioni subiscono importanti deroghe ai successivi commi, rendendo, de facto, il divieto di trattamenti automatizzati non la regola, bensì l'eccezione. Invero, il comma 2 dell'articolo 22 del GDPR prevede tre ipotesi in cui anche un trattamento automatizzato può essere considerato legittimo. In altri termini, il divieto non si applica laddove vi sia un contratto, una legge unionale o nazionale, ovvero nel caso in cui sia reso il consenso del soggetto destinatario di tale trattamento. In particolare, la deroga relativa all'accordo dell'interessato acuisce il rischio del c.d. *blind consent*, ossia la tendenza di prestare il proprio consenso, contrattuale o meno, ai fornitori dei servizi senza effettuare una sostanziale lettura dei termini e delle condizioni del trattamento (Marchetti e Casonato 2021). Tale tendenza diffusa deriva dal senso di urgenza di ottenere un rapido accesso ai servizi. Quindi, la previsione del GDPR che enuncia la necessità di un controllo umano è lacunosa per due ordini di ragioni: uno quantitativo e uno qualitativo. In primis, l'ambito applicativo riguarda solamente i dati personali e ciò esclude un gran numero di ipotesi in cui l'IA potrebbe operare; in secundis, le deroghe di cui al comma 2 dell'articolo 22 indeboliscono ulteriormente tale disposizione. Alla luce di quanto osservato, appare fondata la visione secondo la quale non si può ritenere esistente nel GDPR un diritto sufficientemente solido ed esteso al controllo umano nei trattamenti automatizzati da parte dell'IA.

Il Regolamento europeo sull'Intelligenza artificiale tenta di offrire una soluzione a tali aspetti problematici superando i limiti del GDPR. In particolare, una norma centrale è l'articolo 14 dell'AI Act – rubricato *Human oversight*. L'analisi di tale disposizione non può che cominciare dal primo comma. Infatti, qui si prevede che al fine di garantire la supervisione dell'Uomo sull'IA rientrante nella categoria 'ad alto rischio', quest'ultima dovrà essere ideata, progettata e realizzata in un modo compatibile con le capacità di comprensione e di gestione dell'essere umano. In altre parole, un sistema di IA ad alto rischio dovrà essere *human friendly by design*. La seconda parte di tale articolo fissa lo scopo di tale approccio, ovverosia quello di prevenire o ridurre i rischi per i diritti fondamentali delle persone oggetto del trattamento da parte dell'IA. Precisa la norma che tale valutazione, effettuata ex ante, deve tenere conto dell'uso 'naturale' di tale sistema, o al più di quello che si può ragionevolmente prevedere in base alle sue caratteristiche[17].

Tale precisazione è foriera di conseguenze, poiché richiama il tema della trasparenza e della spiegabilità: per riuscire ad anticipare un uso ragionevole del sistema occorre conoscere e comprendere il funzionamento dello stesso. In assenza di un sufficiente livello di tali caratteristiche, l'articolo 14 non si può essere rispettare. Già queste prime considerazioni sull'articolo rivelano l'intima connessione tra la *transparency*, la *explainability* e lo *human oversight*: essi sono tutti *conditio sine qua non*. Focalizzando l'attenzione sulla c.d. *explainability*, stante l'imponente sviluppo dell'IA, specialmente i sistemi più avanzati presentano una caratteristica unica e al contempo problematica, ossia l'essere una scatola nera (anche detta *black box*) (Pasquale 2015). In dottrina si discute del c.d. *black box problem*, il quale avrebbe origine da almeno due elementi: in primis, i Big Data che, a causa della loro stessa struttura, non consentono di ripercorrere il percorso decisionale della macchina; in secundis, l'apprendimento automatico – potenziato dalla tecnica del *deep learning* – che restituisce un sistema che prescindere da logiche deterministiche e quindi prevedibili. Dunque, la scarsa capacità esplicativa e l'imprevedibilità delle decisioni rendono l'IA portatrice di una nuova era di incertezza e di ignoto tecnologico (De Mari Casareto Dal Verme 2024, 3).

Questo significa che i passaggi logici che l'IA effettua, una volta che una determinata informazione è stata introiettata nel suo dataset, potrebbero risultare incomprensibili anche a un osservatore esperto. In altre parole, vi sarebbe una sensibile lacuna conoscitiva su quanto avviene nell'arco di tempo compreso tra l'input e l'output del sistema. Dunque, si può comprendere il contenuto dei dati di partenza e di quelli finali, tuttavia rimane oscuro il 'come' l'IA sia giunta a un certo risultato. In particolare, rimangono pressocché sconosciute le motivazioni che hanno indotto il sistema a modificare-manipolare i dati e le ragioni che hanno portato a un determinato esito (Pasceri 2023, 35).

17 Questo è il concetto di *reasonably foreseeable misuse*, il quale viene più volte valorizzato nell'AI Act all'articolo 3, n. 13.

Proseguendo, il comma 3 affronta il tema dell'intensità del controllo umano sul funzionamento del sistema di IA cercando di individuare un concreto contemperamento tra efficienza e diritti. La profondità della sorveglianza umana dipende dai diritti in discussione, dai rischi, dal livello di autonomia del sistema di IA e, infine, dal 'contesto' di riferimento e di funzionamento del sistema. Quindi, l'effettiva applicazione di tale regola deve tenere in considerazione tali quattro caratteristiche, in cui al mutare di una sola di esse si potrebbero verificare importanti scompensi imponendo la ricerca di nuovi equilibri. Il tema potrebbe essere concretizzato con un esempio. Si ipotizzi di voler applicare lo standard dello *human in command* alla luce di un diritto fondamentale in gioco dinanzi a un sistema di IA capace di porlo a un rischio elevato – rectius: con un grado di autonomia elevato – ma all'interno di un ambiente scientifico-sperimentale – c.d. *sandboxes*[18]. Quest'ultimo fattore risulta decisivo nel considerare non necessaria l'applicazione dell'HIC, poiché non vi sarebbe un 'reale' pericolo per il diritto.

Da ultimo, è opportuno osservare il comma 4. Quest'ultimo elenca le effettive misure che dovrebbero essere offerte all'utilizzatore finale per concretizzare il diritto al controllo umano sull'IA. Queste sono principalmente quelle legate alla possibilità di discostarsi dall'output prodotto dal sistema nella consapevolezza che lo stesso può essere errato, o comunque impreciso, alla luce dell'imperfezione del processo decisionale. Tale consapevolezza è cruciale e sarà uno dei punti più problematici nell'uso dei sistemi di IA ad alto rischio, poiché il rischio di accettazione acritica della decisione umana su quanto 'suggerito' dal sistema è concreto. In dottrina si è parlato di un effetto c.d. 'pecorone' (*anchoring effect*) (Garapon e Lassègue 2018, 239; Simoncini 2019, 81-83), ossia l'incapacità dell'Uomo di analizzare in maniera critica la decisione dell'IA in forza del mito dell'infallibilità della stessa. In altre parole, l'opzione individuata dal sistema diverrebbe la *default-option force* (Simoncini 2020, 39)[19], la quale godrebbe di un plusvalore di significato. Ovverosia, salvo palesi errori, per discostarsene servirebbe un intenso sforzo dell'essere umano che appare improbabile alla luce della forza persuasiva-psicologica del sistema. L'output dell'IA assume un simile valore poiché, dopotutto, riesce ad alleggerire il c.d. *burden of motivation*. Questa tematica sarà fondamentale perché solamente un vaglio 'umano' della decisione renderà quest'ultima 'accettabile-giustificabile' agli occhi dei consociati. Accanto alla possibilità di discostarsi dalla decisione dell'IA, l'AI Act aggiunge l'obbligo di inserire un comando d'emergenza che riesca a fermare l'attività del sistema di IA, ovvero di consentire l'intervento immediato dell'essere umano. Questo '*stop button*' ha sollevato polemiche da parte dell'industria che non lo giudica con favore.

Osservando l'impianto dell'articolo 14 dell'AI Act, si comprende l'ampiezza potenziale della sua applicazione, poiché i sistemi ad alto rischio occupano una percentuale importante dei sistemi di IA. Il rischio di un'interpretazione eccessivamente rigida potrebbe comportare alcuni problemi e il richiamo del comma 3 alle 'circostanze' di utilizzo dell'IA potrebbe non essere sufficiente per un 'equo' contemperamento tra le esigenze di tutela dei diritti e le necessità di sviluppo tecnologico e di competitività dell'UE. A tali dubbi daranno risposte gli interpreti, in particolare i giudici e le autorità (europee e nazionali) deputate al controllo circa l'attuazione del Regolamento. In conclusione, i nuovi diritti fondamentali che dovranno essere riconosciuti all'interno del sistema costituzionale italiano per consentire un governo 'umano-centrico' dell'IA sono rinvenibili sia nella legislazione europea sia nella legislazione nazionale. In tale capitolo ne abbiamo intravisti almeno tre, ossia *AI transparency*, *explainability* e *human first*.

Tali diritti fanno parte di una costellazione di principi intimamente legati tra di loro. Servirà una continua e proficua interazione tra gli stessi per consentire all'Europa di ergersi a 'faro mondiale', capace di concentrare i valori occidentali volti a un 'giusto' equilibrio tra l'ambizione della *téchne* e la tutela dell'Uomo. Tale contemperamento non può

18 L'articolo 3, n.55, dell'AI Act definisce la c.d. *sandbox* nei seguenti termini: "'*AI regulatory sandbox' means a controlled framework set up by a competent authority which offers providers or prospective providers of AI systems the possibility to develop, train, validate and test, where appropriate in real-world conditions, an innovative AI system, pursuant to a sandbox plan for a limited time under regulatory supervision*". In un simile contesto potrebbe essere derogata la normativa vigente al fine di consentire una determinata ricerca scientifica. A titolo esemplificativo, si veda il *Considerando* n.138 dell'AI Act.

19 Questa categoria deriva dai fondamenti della *nudging theory*. Si rimanda a Thaler e Sunstein (2021).

che essere frutto dell'intervento del diritto nella sua dimensione astratta, ma si evidenzia la necessità che le norme si trasformino da *law (right) in the books* in *law (right) in action* (Pound 1910). In caso contrario, non si potrà che ammettere l'abdicazione assoluta dello ius dal suo ruolo di co-protagonista – a fianco alla tecnica – e la sua riduzione a mera comparsa nella storia dell'umanità.

3. IA e Pubblica amministrazione

Alla luce dell'analisi svolta circa la genesi dei nuovi diritti in relazione all'avvento dell'IA, risulta opportuno cominciare a osservare gli effetti di una sua applicazione nel contesto della PA italiana. Questa scelta deriva dalla necessità di declinare il presente lavoro anche verso un ambito importante alla luce degli attori e dei diritti coinvolti.

Senza ombra di dubbio, la PA rappresenta un settore di grande interesse, ove l'applicazione dei sistemi di IA costituisce una delle tematiche più promettenti e, al contempo, complesse. La ratio di dedicare una particolare attenzione alla Pubblica amministrazione deriva dalla circostanza per cui, in tale ambito, il rapporto individuo-pubblico, e quindi la soggezione del cittadino a un potere statale, è particolarmente evidente.

In tempi recenti, proprio al procedimento amministrativo, dottrina e giurisprudenza hanno dedicato sempre maggiori riflessioni, nel tentativo di individuare un contemperamento tra la spinta dell'IA e la tutela dei diritti delle persone. Infatti, l'impiego di strumenti-sistemi di IA nella fase preparatoria di un provvedimento amministrativo e nella sua effettiva emanazione assume dei connotati critici che necessitano di attenta ponderazione e riflessione, poiché vi potrebbero essere conseguenze rilevanti per i cittadini. In altre parole, l'automazione (anche solo parziale) della decisione finale e della sua fase prodromico-valutativa erode l'idea di un potere pubblico 'antropocentrico' e può incrinare il necessario rapporto fiduciario che lega la PA al cittadino e che giustifica l'accettazione del privato di una decisione che lo riguardi. L'applicazione dei sistemi di IA in questo campo sta subendo un lento, ma costante incremento. Di conseguenza, sono sorte controversie circa la correttezza delle decisioni e delle modalità di impiego di tale tecnologia, le quali sono giunte sino al Consiglio di Stato. La sede giurisdizionale consente di assumere uno sguardo privilegiato, poiché facilita l'estrapolazione di quei principi-diritti che dovranno trovare concretizzazione nei futuri progetti.

Nel contesto italiano vi sono alcuni casi rilevanti in cui sono stati affrontati e analizzati gli algoritmi e il loro utilizzo all'interno della Pubblica amministrazione. Il primo è quello che riguarda la riforma della c.d. Buona scuola in forza della legge n. 107/2015.

In tale occasione, la sezione III-bis del TAR Lazio, con la sentenza n. 3769/2017, ha avuto l'occasione di affrontare la tematica dell'impiego degli algoritmi e delle procedure automatizzate all'interno della PA, avviando una riflessione in seno alla giurisprudenza amministrativa che ha subìto molteplici variazioni ed evoluzioni nel tempo. Questa pronuncia riconobbe che l'algoritmo, decidendo, de facto, le destinazioni degli insegnanti, si integra nel contesto di un procedimento amministrativo e, conseguentemente, dà origine al diritto all'accesso al codice sorgente e alle sottese logiche di funzionamento dello stesso ai sensi della legge n. 241/1990. Dunque si giunse alla conclusione che, alla luce dell'interesse pubblico in gioco, l'argomento della riservatezza dell'algoritmo, in quanto 'opera d'ingegno', non avrebbe potuto trovare accoglimento, non potendosi considerare come una causa legittima di esclusione ai sensi dell'articolo 24 della medesima legge. Inoltre, la circostanza che il Ministero avesse consegnato un documento che 'traduceva' in lingua italiana il linguaggio del programma, non era sufficiente ad escludere l'interesse a accedere direttamente a tale codice sorgente per poter verificare che la traduzione fosse effettivamente corretta. Solamente con tale modalità, infatti, è possibile controllare il funzionamento del programma e il suo rispetto della normativa vigente. Purtuttavia, e ben più rilevante ai fini del presente lavoro, il Tribunale – anche se solo *in obiter* – ha affermato che l'uso di un algoritmo decisionale all'interno di un procedimento amministrativo sarebbe stato valido e legittimo solamente a condizione che si fosse in un caso in cui fosse emerso l'esercizio di un c.d. potere amministrativo vincolato. Al contrario, laddove il contesto fosse stato quello di un c.d. potere amministrativo discrezionale, la conclusione sarebbe stata radicalmente opposta, ossia in senso negativo. Questa statuizione, benché *in obiter*, ha assunto crescente valore: invero alcune successive pronunce hanno ripreso tali argomentazioni trasformandole *in ratio decidendi,* costituendo così un terreno fertile per la nascita e consolidamento di un orientamento giurisprudenziale.

Le successive sentenze n. 9224/2018 e n. 9230/2018 del suddetto TAR Lazio hanno ribadito con vigore la necessaria natura servente dello strumento algoritmico, dovendo l'Uomo mantenere il controllo dei sistemi-procedure informatizzate e automatizzate. In questa prospettiva, un sistema di IA può al più divenire strumento ausiliario nelle mani del funzionario pubblico, dovendosi sempre garantire una qualche forma di 'dominio' sul procedimento amministrativo in ossequio agli articoli 7, 8, 10 della legge n. 241/1990 e 3, 24 e 97 della Costituzione. Dunque, in base a queste direttive interpretative, e anche nel contesto di una attività amministrativa vincolata, si può concludere che non vi può essere una sostituzione del contributo umano con un sistema automatizzato-artificiale, anche se si può arrivare a sostenere che più l'attività della PA è vincolata, minore deve essere il 'controllo umano'. Ciò senza, tuttavia, mai superare la *red line*, rappresentata dalla necessità di mantenere sempre l'Uomo come dominus.

Si comprende, a questo punto, l'atteggiamento di tale primo orientamento di apparente chiusura rispetto ai sistemi di IA automatizzati, poiché il cuore del procedimento amministrativo dovrebbe consistere nell'interazione e interlocuzione tra essere umani. In altri termini, benché la tecnologia possa 'servire' l'Uomo, si afferma il necessario 'volto umano' della PA.

Appare lapalissiano come tale impostazione teorica ricerchi di evitare un'automatizzazione della procedura amministrativa con conseguente spersonalizzazione della fase decisoria. Riprendendo la categorizzazione accennata nel paragrafo 2, parrebbe che tale orientamento non ammetta l'accesso all'HOTL e all'HOUTL, poiché richiede una prevalenza del contributo 'umano' e ciò già nel primo modello potrebbe essere mancante.

In tal guisa, si introduce un primordiale principio di non esclusività dell'attività 'artificiale' del sistema di IA, ovverosia il c.d. *human first* (benché limitato all'*Human in command* e all'*Human in the loop*), che è stato poi sancito definitivamente e con una portata 'maggiore' dall'articolo 14 dell'AI Act. Purtuttavia, benché vi siano degli interessanti spunti, questa prima riflessione dei giudici amministrativi poggia tralatiziamente sulla distinzione tra potere amministrativo vincolato e discrezionale. Quest'ultima però, come si dirà, non consente di impostare correttamente la riflessione, ovverosia di inglobare nel contemperamento dei principi-valori costituzionali anche quelli dell'efficienza, efficacia e buon funzionamento della Pubblica amministrazione. Questa conclusione, benché con l'intento di 'proteggere' l'Uomo, esclude aprioristicamente una riflessione proficua sull'uso dell'IA in un'ampia serie di ipotesi in cui, invece, ciò potrebbe avvenire sempre nel rispetto dei diritti delle persone.

La giurisprudenza amministrativa più recente ha ripreso i primi approdi dei giudici di primo grado chiarendo alcuni aspetti e ripensando alcune categorie. Un primo intervento dei giudici di Palazzo Spada[20], può essere rinvenuto nella sentenza della sezione VI del Consiglio di Stato, n. 2270/2019. In tale occasione, i giudici pur non superando l'asserita utilità della distinzione tra attività amministrativa vincolata e discrezionale, si sono concentrati su un altro aspetto, ossia l'auspicabile progressivo utilizzo dei sistemi di IA nella PA italiana.

L'impostazione iniziale nelle sentenze poc'anzi riportate aveva tentato di comprimere l'utilizzo dei sistemi automatizzati decisionali nel quadro delle attività amministrative vincolate. Dunque, pur ammettendo l'impossibilità di bloccare l'ingresso dell'IA nella PA, i giudici delle prime cure – ossia il primo grado innanzi al TAR – non avevano valorizzato appieno l'utilità della tecnologia, ammettendola solo in chiave residuale e in funzione esclusivamente ausiliare. Invece, il Consiglio di Stato, in questo primo intervento, si è espresso con favore all'introduzione di processi automatizzati all'interno del procedimento amministrativo, poiché ciò avrebbe anche aumentato l'efficacia e il buon funzionamento della PA in ossequio all'articolo 97 della Costituzione e all'articolo 1 della legge n. 241/1990. Inoltre, un simile *favor* nei confronti della tecnologia sarebbe stato conforme anche al Codice dell'Amministrazione digitale (D.Lgs. n. 82/2005) e agli influssi unionali. In altri termini, e specialmente laddove si fosse dinanzi a un gran numero di attività seriali-standardizzate in cui non fossero necessarie valutazioni discrezionali, l'implementazione di sistemi che consentano una gestione automatica avrebbe dovuto essere auspicabile e opportuna per aumentare la velocità dei procedimenti e per ridurre i casi di errori 'umani' dovuti a negligenza – oltre a una superiore garanzia circa l'imparzialità della decisione finale.

20 Con questa espressione si fa riferimento al Consiglio di Stato.

Alla luce di tali argomentazioni, si può osservare come i giudici abbiano inteso incentivare una declinazione della PA in chiave di e-government. Purtuttavia, va specificato che il Consiglio di Stato ha comunque ribadito la necessità che il funzionamento del sistema deve essere assolutamente conoscibile, sia al privato sia al giudice, in un eventuale giudizio. Questo tramite una traduzione chiara, completa ed effettiva del funzionamento del sistema. Dunque, la sentenza n. 2270/2019 del Consiglio di Stato, pur esprimendosi favorevolmente all'utilizzo di sistemi decisionali automatizzati in quanto coerenti con una PA che accoglie l'evoluzione tecnologica, ha mantenuto ferma la distinzione tra la categoria delle attività vincolate e quelle discrezionali. Cionondimeno, questo approdo del Consiglio di Stato non è quello definitivo. Infatti, è individuabile un'ulteriore evoluzione all'interno della sezione VI di Palazzo Spada con la sentenza n. 8472/2019. In tale occasione, i giudici hanno affermato che la suddetta distinzione non parrebbe più essere così attuale non potendosi, quindi, ritenere che l'utilizzo di tecniche automatizzate sia circoscritto alle sole attività amministrative vincolate. Invero, anche nelle attività connotate da discrezionalità potrebbe essere auspicabile l'introduzione di moduli organizzativi che utilizzino le potenzialità della c.d. Rivoluzione 4.0. Queste considerazioni nascono dall'idea secondo la quale l'elemento rilevante non è l'ambito di utilizzo degli algoritmi, bensì le cautele e i principi giuridici che trovano applicazione in chiave protettiva. Dunque, ciò a cui si deve guardare sarebbe il rispetto e la compatibilità di tali procedure automatizzate con le direttive generali del sistema giuridico. In ogni caso, benché in tale occasione si sia ammesso l'utilizzo degli algoritmi anche nel seno di un'attività a carattere discrezionale della PA, questo non implica una c.d. 'delega in bianco' all'amministrazione, dovendo essa sempre rispettare alcuni principi e regole fondamentali.

Purtuttavia, come ha segnalato autorevole dottrina, tale apertura alle c.d. attività amministrative discrezionali dovrebbe intendersi circoscritta tendenzialmente alle sole ipotesi di c.d. discrezionalità tecnica e non, invece, ai casi di c.d. discrezionalità 'pura'. In queste ultime ipotesi, le variabili e le circostanze da tenere in considerazione sono molteplici in base al caso concreto tanto che la stessa legge non regolamenta in toto tali circostanze lasciando il potere alla PA di valutare le singole situazioni. Per tali motivi, l'apertura giurisprudenziale dovrebbe essere limitata ai casi in cui troverebbe applicazione un basso grado di 'discrezionalità'.

Lo stesso Regolamento sull'IA con i suoi principi di trasparenza, spiegabilità e di *human oversight* parrebbe volgere in tale direzione imponendo un certo grado di controllo umano sull'attività dell'IA che aumenta in modo proporzionale in base alla rilevanza dei diritti in gioco.

Indubbiamente, come già anticipato, i giudici hanno ribadito l'opportunità di accogliere anche in seno alla PA la c.d. rivoluzione digitale, ma nel rispetto di alcuni principi fondamentali dell'ordinamento. In questa pronuncia, autorevole dottrina ha individuato l'enunciazione puntuale del principio di legalità algoritmica attraverso l'interpretazione del GDPR. Questa 'nuova' categoria impone che l'utilizzo da parte della PA di tali tecnologie avvenga nel rispetto dei generali principi costituzionali, europei e di quelli specifici del diritto amministrativo. In altre parole, sono le *conditio sine qua non* per delle decisioni automatizzate legittime da parte della PA.

Come si dirà di seguito, i principi individuati in tale occasione sono stati ripresi, rinforzati e arricchiti anche dall'AI Act, poiché il solo GDPR, come già evidenziato, non era sufficiente a garantire una piena tutela, a causa del suo ristretto ambito applicativo e delle numerose eccezioni che lo stesso prevede. I giudici, oltre che considerare gli articoli 13 comma 2 lettera f), 14 comma 2 lettera g), 15 comma 1 lettera h) e nell'articolo 22 comma 1, si soffermano anche sull'articolo 41 – rubricato "*Diritto ad una buona amministrazione*" –, comma 1 e 2, della Carta dei Diritti fondamentali dell'Unione europea (CDFUE).

In virtù di tale previsione, il principio di trasparenza algoritmica si declinerebbe nel senso che la PA, nel momento in cui si accinge a prendere una decisione che potrebbe incidere su un diritto di una persona, deve instaurare una interlocuzione (umana) prima di agire, con la conseguente necessità di mettere a disposizione le proprie informazioni e di motivare puntualmente le ragioni della propria determinazione.

Quindi, già da queste considerazioni si può intuire come la giurisprudenza, con la sua attività ermeneutica, abbia tentato e tenti di 'guidare' la transizione della PA nel quadro della rivoluzione tecnologica dell'IA. Tuttavia, con la sentenza n. 8472/2019 e con gli annessi richiami al GDPR, il Consiglio di Stato non si è limitato a un'analisi del principio di trasparenza, bensì ha tentato di andare oltre. In estrema sintesi, i giudici

si sono orientati al 'governo dell'algoritmo' inteso tout court, il quale deve essere, oltre che trasparente, anche sorvegliato dall'Uomo e non discriminatorio.

Circa la sorveglianza umana – ad oggi lo *human oversight* dell'articolo 14 dell'AI Act – i giudici di Palazzo Spada, riprendendo il modello dello *human in the loop* (con tutte le sue possibili varianti già viste), hanno affermato la necessità che il sistema interagisca proficuamente con l'essere umano non potendosi accettare procedimenti amministrativi in cui il dominus sia la macchina e non l'Uomo. Invece, per quanto concerne la non discriminazione algoritmica (i c.d. *algorithmic bias*), il Consiglio di Stato ha richiamato in particolare il *Considerando* 71 del GDPR. Una parte degli interpreti, in forza di tale riferimento, benché i giudici non avessero approfondito tale profilo, ha derivato anche il diritto al fatto che i dati utilizzati nella decisione siano corretti e di congrua 'qualità'. Purtuttavia, come si ha avuto modo di evidenziare, tale Regolamento – a causa del suo ambito applicativo, delle sue eccezioni e del valore giuridico dei *Considerando* nel quadro normativo europeo – non riusciva a garantire una piena e stabile tutela dinanzi ai sistemi di Intelligenza artificiale. Questo vulnus parrebbe essere stato colmato con l'AI Act, ma solo la giurisprudenza tramite la sua attività ermeneutica potrà dare risposte più precise su tali questioni.

In conclusione, dall'analisi di questa pronuncia del Consiglio di Stato, si può affermare che la giurisprudenza amministrativa abbia ricoperto un ruolo propulsivo non indifferente per un efficace contemperamento: da un lato, le esigenze di sviluppo e implementazione dell'IA in ossequio ai principi di efficacia, buon andamento ed economicità dell'azione della PA; dall'altro lato, quella relativa alla necessità che sia garantita la tutela dei diritti fondamentali delle persone sanciti da norme interne e unionali. Inoltre, tale orientamento appare ad oggi consolidato, poiché successivamente i giudici amministrativi sono tornati su tali argomenti specificando e ampliando alcuni punti e tematiche. Infatti, si può individuare almeno un altro significativo arresto giurisprudenziale, la sentenza del TAR Napoli, sez. III, n. 7003/2022. In tale occasione, i giudicanti hanno ribadito alcuni aspetti fondamentali riprendendo le due sentenze poc'anzi viste del Consiglio di Stato del 2019. In primis, il ricorso a una decisione c.d. algoritmica adottata in seno alla PA sarebbe auspicabile, in quanto idonea ad accrescere l'efficienza e il buon funzionamento dell'amministrazione pubblica. Inoltre, tale tecnologia non sarebbe applicabile solamente nel contesto di un'attività vincolata, bensì anche di una discrezionale. Ciò a condizione che sia garantito un certo grado di controllo umano – il TAR Napoli richiama la nozione di *Human in the loop* –, una congrua motivazione ed un certo grado di trasparenza. Su tale ultimo aspetto, è significativo come il Tribunale amministrativo regionale partenopeo espanda alcune considerazioni circa il principio di trasparenza in relazione all'uso di decisioni automatizzate della sentenza del Consiglio di Stato n. 8472/2019. Ovvero afferma che nella ricostruzione del 'significato' di un algoritmo e del suo funzionamento occorrono non solo competenze giuridiche, ma anche tecniche (ad esempio, in campo informatico e statistico). Di conseguenza, queste branche del sapere devono operare sinergicamente in modo che una 'regola tecnica' diventi una 'regola giuridica' comprensibile ai giudici e ai cittadini.

Conclusioni

Volendo compiere delle riflessioni finali e di sistema, appare essenziale evidenziare l'importanza dell'introduzione dell'IA in seno alla PA e del suo futuro sviluppo. Invero, tale nuova tecnologia dovrà essere inserita all'interno del contesto amministrativo al fine di favorire una c.d. buona amministrazione in ossequio al dettato costituzionale senza, però, trascurare il rispetto dei diritti fondamentali dei cittadini, anche mediante i citati principi dell'*human first*, della *transparency* e dell'*explainability*. Il volto 'antropocentrico' della PA rimane ad oggi un imperativo categorico incomprimibile, pertanto l'AI Act pare essere il miglior strumento a disposizione per tutelare tali aspirazioni.

Bibliografia

Abbott R. (2020), *The Reasonable Robot. Artificial Intelligence and the Law*, Cambridge, CUP

Bradford A. (2020), *The Brussels Effect: How the European Union Rules the World*, Oxford, OUP

Cal S. (2023), Il quadro normativo vigente in materia di IA nella pubblica amministrazione, in Belisario E., Cassano G. (a cura di), *Intelligenza artificiale per la pubblica amministrazione. Principi e regole del procedimento amministrativo algoritmico*, Pisa, Pacini Editore, cap.3

Casonato C. (2019), Costituzione e intelligenza artificiale: un'agenda per il prossimo futuro, *Rivista di Biodiritto*, n. speciale 2, pp.711-725

Cassano G. (2024), Note minime sul D.D.L. in materia di Intelligenza Artificiale, *Diritto di Internet*, 28, n.3, pp.579-590

De Mari Casareto dal Verme T. (2024), *Intelligenza artificiale e responsabilità civile. Uno studio sui criteri di imputazione*, Collana della Facoltà di Giurisprudenza dell'Università di Trento n.48, Napoli, Editoriale Scientifica

Erie M.S., Streinz T. (2021), The Beijing Effect: China's Digital Silk Road as Transnational Data Governance, *Journal of International Law and Politics*, 54, n.1, pp.1-91

European Commission (2019), *Ethics guidelines for trustworthy AI*, Brussell, European Commission

Finocchiaro G. (2024a), *Intelligenza artificiale. Quali regole?*, Bologna, il Mulino

Finocchiaro G. (2024b), *Diritto dell'intelligenza artificiale*, Bologna, Zanichelli

Frosini V. (1983), L'informatica e la pubblica amministrazione, *Rivista trimestrale di diritto pubblico*, 33, n.2, pp.483-494

Garapon A., Lassègue J. (2018), *Justice digitale. Révolution graphique et rupture anthropologique*, Paris, Presses Universitaires de France

Laviola F. (2020), Algoritmico, troppo algoritmico: decisioni amministrative automatizzate, protezione dei dati personali e tutela delle libertà dei cittadini alla luce della più recente giurisprudenza amministrativa, *Rivista di Biodiritto*, n.3, pp.389-440

Marchetti B., Casonato C. (2021), Prime osservazioni sulla proposta di Regolamento dell'Unione europea in materia di Intelligenza artificiale, *Rivista di Biodiritto*, n.3, pp.415-437

Pasceri G. (2023), Le scienze argomentative tra stereotipi e veri pregiudizi: la 'black box', *Ciberspazio e Diritto*, 24, n.73, pp.23-44

Pasquale F. (2015), *The Black-Box Society: The Secret Algorithms That Control Money and Information*, Cambridge (MA), Harvard University Press

Pietropaoli S. (2020), Fine del diritto? L'intelligenza artificiale e il futuro del giurista, in Dorigo S. (a cura di), *Il ragionamento giuridico nell'era dell'Intelligenza Artificiale*, Pisa, Pacini Giuridica, pp.101-120

Pop A.M. (2024), Il risk-based approach, la governance e le sanzioni nell'AI Act, *Ciberspazio e Diritto*, 25, n.3, pp.423-436

Pound R. (1910), Law in Books and Law in Action, *American Law Review*, n.34, pp.12-36

Quintarelli S., Corea F., Fossa F., Loreggia A., Sapienza S. (2019), AI: profili etici. Una prospettiva etica sull'Intelligenza Artificiale: principi diritti e raccomandazioni, *Rivista di Biodiritto*, 3, pp.183-204

Simoncini A. (2020), Amministrazione digitale algoritmica. Il quadro costituzionale, in Perin R.C., Galetta D.-U. (a cura di), *Il diritto dell'amministrazione pubblica digitale*, Torino, Giappichelli, pp.33-52

Simoncini A. (2019), L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà, *Rivista di Biodiritto*, n.1, pp.63-89

Thaler R.H., Sunstein C.R. (2021), *Nudge. Improving Decisions About Health, Wealth, and Happiness*, London, Penguin Publishing Group

Andrei Mihai Pop
andrei.pop@unimi.it
È dottorando di ricerca presso il Dipartimento di Scienze giuridiche Cesare Beccaria dell'Università degli Studi di Milano. Fra le pubblicazioni recenti si segnala: Il risk-based approach, la governance e le sanzioni nell'AI Act, *Ciberspazio e Diritto*, n.78, 2024.

Enrico Ferraris
enrico.ferraris@pagopa.it
È un avvocato specializzato in diritto delle nuove tecnologie con un focus su protezione dei dati, privacy e Intelligenza artificiale. Attualmente ricopre il ruolo di Responsabile Product privacy & ethics presso PagoPA SpA, dove coordina l'implementazione dei principi di *privacy by design* nei prodotti ed *ethics by design* delle soluzioni basate su IA. È attualmente membro del comitato direttivo del Dottorato di ricerca in Giurisprudenza presso l'Università di Milano ed esperto esterno del Comitato etico per le tecnologie emergenti del Comune di Torino.

Collaborazione uomo-macchina
Implicazioni etico-giuridiche e nuove competenze necessarie

Sara Pautasso
Università di Pisa - Assegnista di ricerca in Filosofia del diritto

Nel contesto industriale contemporaneo, la robotica collaborativa rappresenta una frontiera cruciale nell'evoluzione delle tecnologie di automazione. I cobot, progettati per operare in prossimità e sinergia con gli esseri umani, sollevano interrogativi rilevanti in merito all'etica del lavoro, alla sicurezza e alle trasformazioni delle relazioni socio-professionali. Il presente contributo esamina tre dimensioni critiche dell'interazione uomo-macchina: il ruolo della velocità e del controllo temporale come strumenti di potere e di organizzazione dei flussi di lavoro; la tensione etica tra collaborazione e competizione uomo-macchina; infine, l'impatto sull'autonomia e sulla dignità umana. Attraverso un'analisi delle nuove competenze necessarie, si intende valutare se l'integrazione dei cobot nei contesti produttivi possa realmente promuovere valori di equità e giustizia o se, al contrario, possa generare nuove forme di subordinazione.

In the contemporary industrial context, collaborative robotics represents a pivotal frontier in the evolution of automation technologies. Cobots, designed to operate in close proximity and synergy with humans, raise significant questions regarding labour ethics, safety, and the transformation of socio-professional relationships. This paper investigates three critical dimensions of human-machine interaction: the role of speed and temporal control as tools of power and workflow organization; the ethical tension between collaboration and competition, with a focus on how cobots'decision-making capabilities reshape operational dynamics; and finally, the impact on human autonomy and dignity. Through an interdisciplinary analysis, this study aims to assess whether the integration of cobots in production environments can genuinely foster values of equity and justice or, conversely, generate new forms of surveillance and subordination.

DOI: 10.53223/Sinappsi_2025-02-8

Citazione
Pautasso S. (2025), Collaborazione uomo-macchina. Implicazioni etico-giuridiche e nuove competenze necessarie, *Sinappsi*, XV, n.2, pp.101-110

Parole chiave
Automazione
Competenze
Robot

Keywords
Automation
Skills
Robot

Introduzione

Nel contesto industriale contemporaneo, le aziende stanno investendo sempre più nelle tecnologie di automazione e nella robotica, con l'obiettivo di sviluppare sistemi versatili, capaci di operare in ambienti che vanno oltre la sola produzione industriale. Un esempio significativo di questa evoluzione è rappresentato dai cobot (*collaborative*

robots), progettati per interagire dinamicamente con gli esseri umani, offrendo vantaggi in termini di velocità, efficienza e sicurezza, e introducendo nuovi valori associati alla robotica collaborativa nei luoghi di lavoro (*Human-Robot Collaboration*, HRC). Progettati per operare in sinergia con l'uomo, i cobot non rappresentano solo un progresso tecnologico, ma costituiscono anche un'opportunità di riflessione sul futuro della nostra coesistenza con le macchine. L'adozione della robotica collaborativa nelle Piccole e medie imprese (PMI) comporta la sfida di integrare tali tecnologie nei flussi di lavoro esistenti, richiedendo, da un lato, un'adeguata formazione dei dipendenti e, dall'altro, la creazione di nuove procedure operative in grado di bilanciare l'implementazione tecnologica con la capacità di adattamento umano. I cobot, pertanto, non influenzano unicamente le pratiche produttive, ma incidono profondamente anche sugli aspetti relazionali tra i lavoratori. Questo solleva la questione se tali strumenti siano realmente in grado di promuovere valori di equità e giustizia o se, al contrario, contribuiscano a veicolare nuove forme di sorveglianza, oppressione e sfruttamento.

Il presente contributo si concentra su tre aspetti cruciali della collaborazione tra uomo e macchina. In primo luogo, verranno analizzate le implicazioni etiche legate alle scelte di programmazione e i concreti contesti applicativi dei robot (par. 1). Il secondo passaggio argomentativo si concentrerà sulla risoluzione delle questioni etiche emergenti nei processi di collaborazione tra uomo e macchina, esplorando in particolare la tensione tra scenari di collaborazione e di competizione. Sebbene problematiche analoghe siano emerse storicamente, oggi ci troviamo di fronte a una trasformazione significativa: i compiti svolti dai sistemi robotici non si limitano più a mere azioni manuali, ma richiedono vere e proprie capacità di giudizio (par. 2). Verrà dunque indagata la relazione di interdipendenza tra dimensione tecnica e sociale, al fine di valutare le trasformazioni delle relazioni tra dipendenti nei luoghi di lavoro e le risorse formative necessarie per facilitare l'integrazione tecnologica. Infine, si affronterà una questione cruciale e al contempo provocatoria: nell'interazione con i cobot, l'agente umano mantiene un ruolo guida e di piena responsabilità sugli aspetti decisionali, oppure finisce per ritrovarsi in una posizione marginale, subordinata e passiva? Questo interrogativo verrà esaminato attraverso una prospettiva multidimensionale, considerando le implicazioni etiche e operative che emergono dall'integrazione dei cobot nelle pratiche lavorative (par. 3). Ecco, allora, alcune questioni di estrema urgenza a cui cercheremo di rispondere in questo saggio: non rischiamo che all'uomo rimanga solo uno spazio interstiziale per esercitare la propria capacità decisionale? Quali possono essere gli effetti della robotica sulle persone e sui gruppi sociali più vulnerabili? Inoltre, se prevalgono atteggiamenti di automazione dei processi, le competenze dei professionisti non si riducono drasticamente in vista di comportamenti conformistici e omologanti? Infine, che ne è dell'autonomia e della dignità degli esseri umani? Questi interrogativi riguardano i valori etici fondamentali che si originano nell'incontro tra macchine e persone, e che il presente contributo intende ricercare.

1. Le scelte di programmazione e i concreti contesti applicativi dei robot

Il primo aspetto da evidenziare riguarda la definizione di cobot. Il termine cobot nasce dalla sintesi di *collaborative robot* e si riferisce a quegli strumenti all'avanguardia progettati per affiancare in modo dinamico gli esseri umani nello svolgimento di compiti differenti. La prima definizione di cobot risale al 1996 ad opera di Colgate, Wannasuphoprasit e Peshkin (1996, 433):

> *A cobot is a robotic device which manipulates objects in collaboration with a human operator. A cobot provides assistance to the human operator by setting up virtual surfaces which can be used to constrain and guide motion. While conventional servo-actuated haptic displays may be used in this way also, an important distinction is that, while haptic displays are active devices which can supply energy to the human operator, cobot are intrinsically passive. This is because cobot do not use servos to implement constraint, but instead employ "steerable" nonholonomic joints. As a consequence of their passivity, cobot are potentially well-suited to safety-critical tasks (e.g. surgery) or those which involve large interaction forces (e.g. automobile assembly)*.

Questa definizione ha il merito di mettere in luce la differenza con gli strumenti robotici industriali tradizionali, progettati per operare separatamente dagli esseri umani e privi di capacità di adattamento

in tempo reale alle variazioni dell'ambiente. Peraltro, i robot industriali richiedono ampi spazi per la loro struttura e per il raggio di movimento, mentre i cobot sono notevolmente più compatti e versatili. Caratteristiche queste ultime che, non solo accomunano la variegata famiglia di modelli di cobot presenti sul mercato, ma li rendono particolarmente adatti per operare in sinergia con le persone. Non sembra allora inappropriata la definizione dei cobot come "oggetti di prossimità"[1] (Isceri e Macioce 2022, 25). L'aspetto di prossimità che li caratterizza trasforma il modo in cui i lavoratori si rapportano alle macchine; non è solo una differenza semantica, ma pratica e concreta. L'unione in tempo reale dell'uomo e del robot in uno spazio di lavoro condiviso colloca il cobot nella posizione di partner con il quale l'essere umano deve collaborare in modo analogo agli altri colleghi. Il carattere eterogeneo di questa classe di sistemi emerge chiaramente se si osserva e si confronta l'autonomia di un braccio robotico industriale con l'autonomia di un cobot utilizzato nei grandi supermercati statunitensi, finalizzato al miglioramento degli inventari, alla scansione degli scaffali, alla registrazione degli articoli e all'organizzazione dei magazzini. Tale evoluzione contrasta nettamente con la tradizionale separazione tra robot e lavoratore umano, una separazione consolidata dai numerosi protocolli operativi e requisiti di sicurezza. La percezione negativa dei robot come sistemi pericolosi sembra quindi venire meno. Infatti, l'introduzione del concetto di collaborazione uomo-robot promuove un nuovo modo di pensare il robot, che modifica gli atteggiamenti preesistenti (Fletcher e Webb 2017). Le loro dimensioni ridotte, la leggerezza e la potenza limitata li rendono prodotti notevolmente differenti dai robot convenzionali. Inoltre, l'implementazione dei cobot è spesso vista dalle aziende come un miglioramento delle condizioni lavorative, liberando parzialmente gli operatori dalle attività ripetitive e permettendo loro di concentrarsi su compiti più qualificanti (*ibidem*). Tuttavia, rimane da chiedersi se questa promessa di miglioramento venga effettivamente mantenuta nella pratica. Infatti, più le tecnologie sono progettate per interagire e collaborare con l'umano, più si rende necessaria una profonda riflessione sui modi in cui formiamo relazioni con gli strumenti tecnologici. Le implicazioni della c.d. Quarta Rivoluzione industriale non riguardano soltanto la realtà aziendale e i processi economici e produttivi in termini di investimenti, bensì l'organizzazione della vita lavorativa e lo spazio comunicativo e collaborativo con gli agenti 'artificiali'. Secondo Klaus Schwab (2016), tali cambiamenti non si limitano all'aspetto tecnico, ma si caratterizzano per il modo in cui l'ambiente umano si integra e si fonde con quello digitale e tecnologico[2]. Questa fusione evoca uno scenario di radicale trasformazione del nostro modo di comunicare, educare, usufruire di beni e servizi e, nello specifico, lavorare. Pertanto, sarebbe del tutto insufficiente considerare le implicazioni della Quarta Rivoluzione industriale da un punto di vista meramente tecnico[3]. Le nuove realizzazioni tecnologiche sono sempre il frutto di processi sociali complessi e di strutture politiche che ne sostengono lo sviluppo e, per questo, coinvolgono tutti gli ambiti della vita. Per tali ragioni, l'attenzione al cambiamento dei valori umani coinvolti nelle pratiche lavorative con le macchine si rivela un tema centrale della riflessione etica contemporanea. In quello che segue, si rifletterà sull'urgente necessità sia di esaminare le capacità delle macchine, nella misura in cui modificano la vita dell'uomo, sia di affermare quei valori specificamente umani che ci differenziano da esse. Si tratterà, cioè, di chiedersi se questi strumenti siano in grado di coadiuvare l'uomo nella ricerca di valori di equità e giustizia oppure se mantengano in vita forme di oppressione e sfruttamento, tali da limitare, persino distruggere, quelle caratteristiche che chiamiamo propriamente umane.

1 Isceri e Macioce scrivono a tal proposito: "Peraltro, mentre i robot industriali sono stati progettati per essere separati dagli utilizzatori umani, confinati, per ragioni di sicurezza, i cobot perdono di principio questa connotazione come oggetti pericolosi e divengono oggetti di prossimità" (2022, 25).

2 Klaus Schwab (2016) parla di Quarta Rivoluzione industriale per descrivere la rivoluzione in atto a opera dell'Intelligenza artificiale, che segue quelle del motore a vapore, dell'elettricità e dell'informatica, caratterizzata da innovazioni che hanno dato vita a "tecnologie combinate impiegabili nell'ambito fisico, digitale o biologico".

3 Luciano Floridi (2017) chiama questa fase di radicale trasformazione "quarta rivoluzione" per differenziarla dalla Quarta Rivoluzione industriale. La quarta rivoluzione di cui parla Floridi è una rivoluzione dell'essere e della comprensione dell'umano, che coinvolge tutti gli ambiti della vita.

L'attenzione alla collaborazione uomo-macchina, su cui oggi insiste e si concentra gran parte della ricerca filosofica e giuridica, rivela come la linea di demarcazione tra le due entità, uomo e macchina, sia spesso percepita come inversamente proporzionale. Sembrerebbe, cioè, che più tale linea avanza verso la macchina, più viene sottratto all'uomo. La prima questione riguarda allora come tracciare questa linea di confine, le cui implicazioni costituiscono una delle sfide e delle opportunità più significative del nostro tempo. Nella collaborazione uomo-macchina, le capacità del robot e quelle dell'essere umano sono in relazione e si completano a vicenda, facendo sì che l'uno dipenda dall'altro. Da ciò deriva la necessità di confrontare le loro prestazioni. Il rischio che la collaborazione diventi competizione è insito nel presupposto stesso del loro funzionamento, infatti, se il cobot è visto come veloce, l'umano potrebbe sembrare lento; se il cobot è performante, l'umano risulta vulnerabile e fallibile (Wallace 2021). La percezione del lavoratore di dover essere comparato, o di poter essere sostituito dalla macchina, può divenire una minaccia reale per gli operatori, compromettendone non solo la salute fisica e psicologica, ma anche la capacità di lavorare in team e di mantenere relazioni sociali positive. Infatti, gli operatori umani devono adattarsi a un vasto numero di cambiamenti, non soltanto lavorativi, ma anche fisici, psicologici e sociali (Fletcher e Webb 2017). Tali questioni non coinvolgono solo gli aspetti tecnici legati all'automazione o all'ergonomia fisica, ma sollevano interrogativi di natura valoriale e organizzativa più profonda. A tal proposito, Haenlein e Kaplan sostengono la necessità di studiare, anzitutto, quali task e quali decisioni debbano "essere assunti dai sistemi di AI, quali dagli operatori umani e quali in collaborazione" (2019, 9). La loro analisi mette in evidenza la necessità di una valutazione dettagliata della divisione del lavoro tra esseri umani e tecnologie, al fine di ottimizzare la collaborazione tra le due parti. In altre parole, una percezione ingenua dell'innovazione nei contesti lavorativi potrebbe creare resistenze tra i dipendenti e sottostimare una serie di questioni rilevanti nella gestione delle risorse umane.

Si possono, infatti, ipotizzare diversi scenari. Il primo, in cui la macchina assume un ruolo subordinato, lasciando spazio all'*agency* umana; il secondo, in cui la macchina svolge un ruolo interdipendente e complementare, in grado di potenziare le facoltà umane; il terzo, nel quale la macchina svolge un ruolo sostitutivo dell'uomo (Raisch e Krakowski 2021). Nel caso in cui prevalga l'ultimo scenario, potrebbe emergere un atteggiamento di competizione da parte del lavoratore, accompagnato da una perdita di fiducia nell'apparato tecnologico o da un'eccessiva sensibilità verso ogni genere di errore, anche minimo, commesso dalla macchina. Si potrebbe, cioè, generare una contraddizione intollerabile e insanabile, in cui gli operatori percepiscono la macchina come un'invasione indebita del loro campo d'azione. Al contempo, un approccio eccessivamente favorevole, volto a considerare l'introduzione della tecnologia come automatico miglioramento delle condizioni lavorative, potrebbe mettere in luce i limiti di una formazione del dipendente talvolta inesistente (Brynjolfsson e McAfee 2015). Al di là di queste narrazioni, che sono frequentemente caratterizzate da alti livelli di generalizzazione, toni sensazionalistici e, a seconda degli autori, da venature ottimistiche o pessimistiche, l'impatto effettivo delle nuove tecnologie deriva dalle scelte di progettazione e dalle pratiche di impiego nei processi organizzativi e applicativi. In questa prospettiva, assumono grande importanza sia la dimensione dell'*organization,* ovvero delle scelte di progettazione, sia dall'o*rganizing*, cioè delle forme emergenti di relazione uomo-macchina nei concreti processi organizzativi (Blasutig 2022, 180). La prima dimensione, corrispondente a un movimento di tipo top-down, coincide con le scelte di pianificazione con cui vengono formalmente fissate le regole di funzionamento e assegnate le risorse alle diverse componenti organizzative. In sostanza, perché, cosa, dove, quanto e come la macchina entrerà nell'organizzazione. La seconda dimensione, corrispondente a un movimento di tipo bottom-up, è riconducibile al modo in cui prendono forma i processi e le pratiche organizzative nei concreti contesti di azione e di costruzione intersoggettiva dei significati. In termini più espliciti, Salento (2018, 11) identifica tre tipi di decisioni nel processo progettuale: le prime sono "decisioni di concezione e progettazione", che stabiliscono gli obiettivi dell'artefatto tecnologico, le sue funzioni e le modalità di interazione con gli operatori. Vi sono poi le "decisioni di adozione", che riguardano settori, fasi e processi per i quali l'artefatto viene utilizzato. Infine, Salento parla di "decisioni di utilizzo", che riguardano le modalità concrete di impiego delle tecnologie. Dunque, da una parte il movimento top-down definisce il tipo di rapporto

che si viene a creare tra uomo e robot in base alle scelte dei programmatori (il grado di autonomia, le ragioni dell'impiego dell'artefatto tecnologico, le modalità di interazione con l'uomo ecc.), dall'altra, il movimento bottom-up delinea concretamente in che modo prenda forma la divisione del lavoro tra agente artificiale e agente umano. In questi casi, come sostiene Blasutig (2022), le scelte di investimento, adozione e utilizzo discendono da tre grandi logiche, operanti congiuntamente, nonostante pesi diversi a seconda delle circostanze. Anzitutto, troviamo la "logica della convenienza", che si basa principalmente su valutazioni di ordine tecnico ed economico, in termini di produttività. Secondariamente, la "logica dell'appropriatezza", in relazione a standard, a norme e pratiche accettate all'interno del settore. Infine, la "logica della affidabilità", relativa all'esperienza e alla costruzione di significati in relazione ai possibili effetti delle macchine, riscontrati in sede applicativa o soltanto immaginati.

Alla luce di queste considerazioni possiamo domandarci: fino a che punto è eticamente corretto che il progettista influenzi il comportamento del lavoratore per orientarlo verso modalità considerate produttive dall'azienda? Nonostante non siano del tutto prevedibili gli sviluppi della relazione tra uomo e macchina nei concreti processi organizzativi, è importante considerare congiuntamente il doppio movimento dell'*organization* e dell'*organizing.* L'importanza di tenere presenti questi fattori è essenziale per creare un sistema automatizzato che sia realmente in grado di supportare l'aspetto umano. La buona riuscita dell'interazione dell'uomo con gli artefatti tecnologici dipende sia dalle scelte progettuali sia dalle configurazioni organizzative concrete. Gli ingegneri stabiliscono i margini di *agency* dell'uomo e fissano le regole fondamentali che guidano l'operato dei cobot, ovvero strutturano lo spazio di interazione e collaborazione tra uomo e macchina, stabilendo quali compiti debbano essere automatizzati e quali prerogative decisionali debbano competere all'uomo. Tuttavia, le esperienze pratiche scaturite dall'uso quotidiano dei cobot arricchiscono e affinano le decisioni top-down, creando un ciclo virtuoso di miglioramento continuo.

2. Competizione o collaborazione uomo-macchina?

L'integrazione dei cobot nei contesti lavorativi solleva anche questioni critiche legate al benessere degli operatori. La European Agency for Safety & Health at Work (2019) ha identificato diversi fattori di rischio psicosociale connessi all'adattamento delle persone ai ritmi e alle modalità operative imposti dai robot. In particolare, l'Agenzia sottolinea come l'esigenza di mantenere una continuità operativa con robot altamente specializzati ed efficienti possa avere effetti deleteri sulla salute mentale degli operatori. I cobot, imponendo standard di velocità ed efficienza, inducono i lavoratori a modificare il proprio comportamento, rischiando di ridurre l'essere umano a mero esecutore di task. Tale riduzione evidenzia i limiti di una formazione spesso del tutto inesistente e acuisce la competizione con la macchina. La prossimità dei cobot agli esseri umani introduce quindi un cambiamento sostanziale nelle dinamiche socio-culturali ed economico-giuridiche contemporanee. I sistemi automatizzati di monitoraggio consentono alle aziende un controllo pervasivo delle performance dei dipendenti, monitorando in tempo reale la produttività e individuando immediatamente cali o errori (Fletcher e Webb 2017). Questa sorveglianza pervasiva incide profondamente sul comportamento dei lavoratori, influenzandoli sia in modo diretto che indiretto. Come osserva Hong (2015), tale sorveglianza genera una condizione di inter-passività, in cui le scelte individuali si riducono ad accettare o rifiutare decisioni predefinite dalle macchine. L'inter-passività rappresenta l'incapacità dell'operatore di appropriarsi della tecnologia, subendo una situazione non scelta, ma solamente accettata o respinta, senza reale incidenza sul processo produttivo. Questo fenomeno accentua il potere reificante dell'apparato tecnologico, strettamente connesso all'uniformità tra uomo e macchina. Se la reificazione riduce l'essere umano a semplice oggetto, questi diviene assimilabile alle logiche operative delle macchine, perdendo la propria specificità. I sistemi di sorveglianza intensificano tale tendenza, rischiando di uniformare le dinamiche lavorative secondo standard di efficienza che non riconoscono l'unicità del contributo umano. In gioco è, allora, l'autonomia degli operatori, intesa come capacità di prendere decisioni indipendenti, responsabili e libere da influenze esterne. Questo rischio evidenzia la necessità di ridefinire le modalità di interazione uomo-macchina, salvaguardando l'autonomia decisionale e la dignità del lavoro umano.

Per comprendere questo passaggio, è utile fare un passo indietro. Durante le rivoluzioni industriali passate, l'automazione ha mutato profondamente i profili delle professioni, creando nuove opportunità di lavoro e sostituendone altre (Wiener 1966). Istanze analoghe si sono presentate periodicamente nella storia dello sviluppo tecnologico; tuttavia, oggi, la novità distintiva rispetto al passato riguarda i compiti cognitivi che gli agenti artificiali sono ora in grado di svolgere e che, fino a pochi decenni fa, erano di pertinenza esclusiva degli esseri umani. Le questioni etiche sull'autonomia operativa delle macchine si presentarono già durante la metà del XX Secolo. In quegli anni, Norbert Wiener, padre della cibernetica e precursore della robotica, aveva messo a fuoco questo elemento di radicale novità allo scopo di allertare le organizzazioni sindacali sui potenziali rischi e pericoli che avrebbero potuto interessare l'automazione industriale (1966). Come notava Wiener, non si trattava soltanto di una potenziale sostituzione dell'uomo con la macchina nella forza lavoro e nell'energia, ma nella capacità di giudizio e nelle facoltà intellettive. Una novità di tale portata, se non fosse stata a vantaggio dei lavoratori, avrebbe condotto non soltanto al peggioramento delle condizioni lavorative e alla disoccupazione di massa, ma alla competizione con la macchina stessa e, nel peggiore dei casi, alla riduzione dell'essere umano alla macchina. Scrive Wiener (1966, 229): "Ho parlato di macchine, ma non soltanto di macchine che possiedono cervelli di ottone e muscoli di ferro. Allorché le persone umane sono organizzate nel sistema che li impiega non secondo le loro piene facoltà di esseri umani responsabili, ma come altrettanti ingranaggi, leve e connessioni, non ha molta importanza il fatto che la loro materia prima sia costituita da carne e da sangue". Nel momento in cui gli esseri umani sono trattati come componenti intercambiabili di una macchina più grande si perde di vista la loro complessità, indipendentemente dal fatto che interagiscano con robot o con persone in carne e ossa. Se l'uomo si adegua ai movimenti del meccanismo, se cioè perde di vista la propria complessità e agisce seguendo uno schema meccanico, si impoveriscono anche le dinamiche sociali e relazionali. Durante la Prima Rivoluzione industriale, l'avvento delle nuove macchine portò a un aumento senza precedenti della produttività, ma contribuì ad una crescente precarietà degli appartenenti a interi settori legati alle vecchie tecniche di produzione, riducendoli all'emarginazione. Da questa situazione ebbero origine manifestazioni di protesta, anche violenta, come il cosiddetto luddismo (da Ned Ludd, un operaio inglese), che praticarono la distruzione dei telai meccanici per ostacolare il processo di dequalificazione dovuto alla meccanizzazione delle loro funzioni lavorative. La reazione violenta ed estrema dei luddisti spinse gruppi di operai, privati dei mezzi di sopravvivenza e in condizioni di estrema miseria, a distruggere le nuove macchine, considerate la ragione prima e intollerabile della loro degradazione (Hobsbawm 1954, 33-53). Per questo motivo, è di fondamentale importanza partire da un'articolazione normativa per l'impiego di sistemi autonomi, nel tentativo di evitare una riduzione dell'umano a meri processi funzionali. Ma, domandiamoci, in che forme, misure e modalità? I cobot rappresentano un esempio tangibile di come i principi cibernetici di Wiener siano stati implementati nella robotica contemporanea. Progettati per operare in sinergia con gli esseri umani in ambienti condivisi, i cobot non solo costituiscono un progresso tecnologico, ma rappresentano anche una sfida e un'opportunità per riconsiderare il rapporto tra lavoro e tecnologia, invitandoci a riflettere sul futuro della nostra coesistenza con le macchine.

Ecco, allora, che, nel contesto della crescente coesistenza con le macchine, l'adozione dei cobot nelle aziende solleva interrogativi fondamentali riguardo alla preservazione dell'autonomia umana e alla qualità delle relazioni interpersonali. Per evitare che la loro introduzione sia percepita come un'imposizione che limita l'autonomia dei lavoratori, è cruciale prevenire la rarefazione delle relazioni interpersonali, sia nei rapporti orizzontali tra colleghi, sia nei rapporti verticali con i dirigenti di riferimento (Plesner e Husted 2022, 166). Il terreno dell'intersoggettività è fondamentale per creare team di lavoro solidi e comunità di pratiche. Questo stesso terreno alimenta il capitale sociale di un'azienda, promuovendo risorse essenziali quali: "la fiducia interpersonale, la condivisione di valori, il senso di appartenenza, il sostegno reciproco e la motivazione a cooperare e contribuire attivamente all'azione collettiva dell'organizzazione" (Blasutig 2022, 201). Mantenere il corretto equilibrio tra l'adozione delle tecnologie e il rispetto dei principi di autonomia e libertà dell'essere umano implica valorizzare le interazioni personali, essenziali per un ambiente di lavoro coeso e collaborativo. Le tecnologie, infatti,

tendono a uniformare i comportamenti umani ai modelli meccanici, trasformando i lavoratori in ingranaggi di un sistema tecnologico predeterminato. Questo processo facilita il controllo, poiché priva gli individui della loro autonomia e responsabilità morale. Le aziende che assegnano compiti ripetitivi relegano i dipendenti a una condizione puramente meccanica, escludendoli da quel "livello intellettuale indispensabile per varcare i limiti del problema e comprenderne l'importanza" (Wiener 1966, 75). Wiener paragona tale condizione a quella delle formiche: insetti privi di valore individuale, sostituibili e rigidamente vincolati a schemi comportamentali predeterminati. Le formiche, guidate dall'istinto e non dall'intelligenza, incarnano una "camicia di forza fisica" che determina la "camicia di forza mentale" responsabile della rigidità dei loro comportamenti (Wiener 1966, 82). Analogamente, il lavoratore ridotto a una funzione ripetitiva rischia di essere declassato a ingranaggio, perdendo il valore della propria umanità. Il rischio è, dunque, duplice: da un lato, ridurre l'essere umano alle sue prestazioni e capacità significa favorire una competizione svantaggiosa con le macchine; dall'altro, antropomorfizzare eccessivamente le macchine rischia di compromettere la comprensione delle loro reali funzioni e di limitare la capacità umana di gestire efficacemente la relazione uomo-macchina. Sebbene le macchine possano eccellere in precisione e velocità, non possono replicare il valore intrinseco dell'umanità.

Per affrontare queste sfide, è necessario ripartire da un modello etico d'impresa fondato su pratiche aziendali resilienti. Questo modello dovrebbe coniugare vantaggi competitivi sostenibili e la promozione di relazioni significative tra i dipendenti. La creazione di ambienti di lavoro collaborativi, dove le interazioni faccia a faccia siano la norma piuttosto che l'eccezione, è essenziale per migliorare il rendimento aziendale e la coesione sociale interna[4]. Ciò richiede che le politiche aziendali pongano gli aspetti relazionali al centro, costruendo comunità lavorative orientate non solo all'efficienza produttiva, ma anche a valori etici condivisi. In questo senso, i cobot, se integrati con consapevolezza, possono fungere da catalizzatori per la riorganizzazione delle relazioni lavorative, promuovendo una cultura aziendale basata sulla collaborazione e sul rispetto reciproco.

3. Le risorse formative dei dipendenti

Come illustrato nel paragrafo precedente, la collaborazione tra esseri umani e macchine comprende non solo un'azione sinergica sul piano operativo, ma anche un'interazione cognitiva basata su una rappresentazione condivisa della realtà e dei problemi da affrontare. Per ottenere una collaborazione proficua tra umani e macchine, gli esseri umani non soltanto devono operare con le tecnologie, ma devono anche comprenderne le logiche (Fabris 2018). Questo richiede un controllo critico per orientare le azioni verso obiettivi comuni e significativi, poiché, in assenza di tale coinvolgimento, si verifica il rischio di ridurre l'autonomia umana a mera conformità alle direttive della macchina. Per tali ragioni, l'introduzione dei cobot deve essere gestita con attenzione, attraverso una gestione del cambiamento che includa un'analisi approfondita delle competenze formative, degli impatti e delle strategie di adattamento. Infatti, le risorse formative disponibili per i dipendenti spesso non sono sufficienti a gestire le innovazioni all'interno delle organizzazioni: per questo, non è solo importante rafforzare la fiducia degli utenti nei sistemi uomo-robot, ma garantire che l'informazione sia chiara e non fuorviante (Fletcher e Webb 2017). Di fronte a tali sfide, è necessario domandarci: quali obiettivi formativi dovrebbero essere coinvolti nei processi innovativi? E, dunque, quali strutture di pensiero e disposizioni mentali occorrono per sostenere l'innovazione e l'eventuale imprevisto?

Secondo l'indagine *The Future of Jobs*[5], la società è ancora lontana dall'elaborazione di percorsi

4 Secondo uno studio condotto da Lee *et al*. (2010) presso la Harvard Medical School, i risultati scientifici più significativi si ottengono quando i ricercatori collaborano in stretta prossimità fisica, entro la stessa struttura edilizia. In questi casi, gli articoli risultano avere un tasso di citazioni mediamente superiore del 45 % rispetto a quelli prodotti da coautori situati in edifici differenti. Gli autori osservano: "*Despite the positive impact of emerging communication technologies on scientific productivity, physical proximity continues to play a critical role in predicting the impact of scientific research*" (*ibidem*, 4).

5 World Economic Forum (2019). Nel Report, il WEF individua dieci competenze chiave: problem solving complesso, pensiero critico, creatività, gestione delle persone, coordinamento, intelligenza emotiva, capacità decisionale, orientamento al servizio, negoziazione e flessibilità cognitiva. Queste abilità favoriscono l'adattabilità, la collaborazione e l'innovazione in un mercato del lavoro in continua evoluzione.

formativi efficaci, capaci di affrontare le sfide emergenti poste dalla robotica collaborativa. Un elemento cruciale per mitigare gli impatti negativi di questa tecnologia è la predisposizione di contratti aziendali in grado di fronteggiare i nuovi rischi, non solo di natura tecnica, ma anche psicologica. La percezione, da parte dei lavoratori, di poter essere sostituiti dalle macchine a causa di una presunta lentezza, o il timore legato alle proprie prestazioni, può costituire una minaccia reale. Tali paure non solo compromettono la salute individuale, ma possono anche minare la capacità di lavorare in team e di mantenere relazioni sociali positive. Se i dipendenti non ricevono una formazione adeguata, possono sviluppare timori legati al licenziamento, alla sopraffazione tecnologica da parte dei cobot, o una dipendenza disfunzionale dalla macchina, con il rischio di abbandonare il lavoro. Il Report evidenzia, infatti, come il 40% dei lavoratori necessiti di una riqualificazione per affrontare le trasformazioni aziendali e lavorative in atto. In poche parole, è necessario un processo di *reskilling* dei dipendenti[6]. Per questo, risulta essenziale individuare le competenze necessarie per far fronte a un mercato del lavoro in continua evoluzione. Diversi studiosi (Holmes *et al.* 2015) concordano nell'affermare che l'attivazione del pensiero critico sia essenziale per garantire interventi consapevoli, decisioni informate e azioni strutturate. Il 'pensiero critico'[7] consente di evitare pregiudizi e derive disumanizzanti, migliorando la qualità del lavoro e della vita quotidiana. Inoltre, esso è cruciale per lo sviluppo dell'autonomia del dipendente, permettendo agli individui di valutare criticamente le situazioni ed evitare un'accettazione passiva delle stesse. Questo approccio facilita l'adattamento ai cambiamenti e promuove una corretta interazione tra esseri umani e macchine. Occorre, pertanto, immaginare strategie etiche e giuridiche per prevenire tali rischi e garantire la qualità dell'esperienza lavorativa condivisa con gli agenti artificiali. Per questo, risulta fondamentale stabilire con chiarezza quali azioni i cobot possano svolgere autonomamente, quali richiedano il controllo umano e quali possano essere realizzate in collaborazione. Ciò garantirebbe uno sviluppo critico e responsabile della robotica collaborativa nei processi decisionali. La principale sfida per i progettisti di nuove tecnologie, come i cobot, consiste, allora, nel considerare fin dall'inizio l'impatto sulle relazioni umane, adottando misure preventive per minimizzare i rischi e massimizzare i benefici. In quest'ottica, migliorare un cobot non significa solo far avanzare tecnologicamente i processi produttivi, ma anche potenziare l'interazione con l'utente e facilitare la familiarità del lavoratore con la macchina (Tamburrini 2020). Come osserva Batya Friedman (1996): "i valori emergono dagli strumenti che costruiamo e dal modo in cui scegliamo di usarli" (trad. dell'Autrice). Friedman sottolinea le difficoltà nell'articolare e tradurre i valori morali in progetti concreti. È quindi essenziale interrogarsi su quali valori siano implicitamente incorporati nella progettazione dei cobot e su come essi influenzino le dinamiche lavorative[8].

In linea con una prospettiva etica e partecipativa, le decisioni sull'introduzione di tecnologie collaborative nei luoghi di lavoro devono coinvolgere attivamente i lavoratori. Non è sufficiente limitarsi a informarli: è necessario garantire la loro presenza effettiva al 'tavolo delle decisioni', affinché le loro preoccupazioni e necessità siano considerate sin dalle fasi iniziali della progettazione, fino all'attuazione e alla valutazione delle tecnologie stesse. Questo approccio partecipativo assicura che le soluzioni adottate siano effettivamente efficaci e rispondenti ai bisogni reali dei lavoratori. Tuttavia, una partecipazione priva di adeguata preparazione rischia di ridursi a una misura superficiale, utile forse a migliorare l'immagine aziendale ma incapace di generare cambiamenti sostanziali.

6 Il *reskilling* consiste nel formare i dipendenti affinché acquisiscano competenze utili a ricoprire ruoli differenti, avviando un processo di riqualificazione personale e professionale. L'*upskilling,* invece, mira a potenziare le competenze già possedute, arricchendo le abilità nel medesimo ambito lavorativo per favorire una crescita professionale.

7 Il pensiero critico è un processo cognitivo definito come "la capacità di pensare in modo razionale e riflessivo su ciò che crediamo e facciamo" (Facione 2015).

8 Determinare preventivamente il valore morale dell'azione umana in relazione all'uso delle tecnologie è fondamentale per valutare i potenziali rischi dei processi di automazione. Ad esempio, la velocità operativa di un cobot diventa un criterio di produttività che costringe anche l'essere umano a lavorare a una velocità maggiore, introducendo, di fatto, un valore specifico – la velocità – nella progettazione del cobot. L'errore è accettare il presupposto che l'uomo debba conformarsi alla velocità della macchina, finendo per categorizzare e ridurre il soggetto alle abilità che possiede.

Per evitare che la partecipazione sia meramente formale, è cruciale garantire che i lavoratori siano adeguatamente formati e che il loro coinvolgimento sia significativo (Isceri e Macioce 2022, 32). In tale prospettiva, Isceri e Macioce sottolineano il ruolo strategico del sindacato nella mitigazione dei rischi tecnologici. Essi evidenziano come il sindacato non debba limitarsi a una partecipazione formale, ma essere autenticamente integrato nel processo decisionale, partecipando alle fasi di test delle procedure informatizzate. Questo consentirebbe di verificare, prima del rilascio, che le prestazioni del software robotico rispettino gli accordi sottoscritti (*ibidem*). Gli autori propongono inoltre l'istituzione di un'azione sindacale aziendale fondata su dati (*data-driven*), in grado di negoziare accordi specifici che orientino lo sviluppo dei cobot e garantiscano il rispetto delle intese raggiunte. Ripensare lo spazio lavorativo in chiave relazionale significa costruire ambienti più umani, attenti alle esigenze dei singoli e capaci di valorizzare la complessità e la vulnerabilità di ciascuno. In questo contesto, la robotica collaborativa può diventare uno strumento di supporto, capace di sostenere anche i lavoratori più fragili, superando il mito dell'autonomia umana e promuovendo un'etica dell'interdipendenza, fondata sulla consapevolezza della comune fragilità e corporeità che ci definisce come esseri umani.

Conclusioni

In conclusione, la discussione sull'integrazione dei cobot non dovrebbe limitarsi a valutare l'efficacia di queste tecnologie nell'ottimizzazione dei processi produttivi, ma dovrebbe concentrarsi su come rafforzare le relazioni sociali e professionali tra i dipendenti. Tra i principali vantaggi derivanti dall'impiego dei cobot figura la possibilità di sollevare l'essere umano da mansioni ripetitive e pericolose, contribuendo a ridurre le attività alienanti e rischiose e a rendere il lavoro più accessibile e sostenibile. Tuttavia, la progettazione della collaborazione tra esseri umani e cobot dovrebbe anche mirare a liberare il potenziale umano, evitando che le macchine assumano comportamenti assertivi, rigidi e prescrittivi. L'obiettivo è garantire che i cobot non assumano una postura assertiva, bensì si configurino come strumenti capaci di supportare e potenziare le capacità umane, rispettando la diversità delle competenze e delle esigenze individuali.

Da un punto di vista giuridico, l'introduzione dei cobot solleva questioni cruciali legate alla responsabilità, alla sicurezza sul lavoro e alla tutela dei diritti dei lavoratori. L'integrazione tecnologica deve quindi essere accompagnata da un aggiornamento normativo che assicuri la protezione dei lavoratori e la promozione di ambienti di lavoro inclusivi e sicuri. Tuttavia, ciò non è sufficiente. La riqualificazione dei dipendenti e l'educazione etica devono essere garantiti come strumenti fondamentali per preparare i lavoratori a interagire responsabilmente con le tecnologie emergenti. Le nuove competenze richieste, infatti, non si limitano alle abilità tecniche necessarie per interagire con i cobot, ma includono anche competenze trasversali. In risposta alle domande iniziali, si può affermare che i cobot possano effettivamente migliorare l'esperienza lavorativa, a condizione che siano progettati per adattarsi alle diversità dei lavoratori, anziché uniformare i loro compiti. Per raggiungere tale finalità, è cruciale perseguire il benessere dei lavoratori, creando ambienti che incentivino la socializzazione, valorizzino il contributo umano e promuovano una collaborazione sinergica non solo con la tecnologia, ma anche tra colleghi. Queste due dimensioni devono coesistere armoniosamente, senza che l'una escluda l'altra. L'etica del lavoro, pertanto, non è chiamata solo a riflettere sulle differenze e sulle somiglianze tra esseri umani e macchine, ma a interrogarsi su come integrare le tecnologie in modo da generare valore non solo economico, ma anche sociale e umano.

Bibliografia

Blasutig G. (2022), L'intelligenza artificiale nelle organizzazioni e la prospettiva della collaborazione uomo-macchina, *Poliarchie/Polyarchies*, 5, n.2, pp.177-209

Brynjolfsson E., McAfee A. (2015), *La nuova rivoluzione delle macchine. Lavoro e prosperità nell'era della tecnologia trionfante*, Milano, Feltrinelli

Colgate J.E., Peshkin M.A., Wannasuphoprasit W. (1996), Cobot: Robots for Collaboration with Human Operators, in *Proceedings of the International Mechanical Engineering Congress and Exhibition*, Atlanta (GA), DSC-Vol. 58, 17-22 novembre, pp.433-439

European Agency for Safety and Health at Work (2019), *OSH and the Future of Work: benefits and risks of artificial intelligence tools in workplaces,* Bilbao, European Agency for Safety and Health at Work

Facione P.A. (2015), *Critical Thinking: What It Is and Why It Counts*, Millbrae, Insight Assessment

Fabris A. (2018), *Etica per le tecnologie dell'informazione e della comunicazione*, Roma, Carocci

Fletcher S.R., Webb P. (2017), Industrial robot ethics: The challenges of closer human collaboration in future manufacturing systems, in Robasto D. (a cura di), *Robot e cobot nell'impresa e nella scuola*, Milano, Franco Angeli, pp.159-169

Floridi L. (2017), *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, Milano, Raffaello Cortina Editore

Friedman B. (1996), Value-sensitive design, *Interactions*, 3, n.6, pp.16-23

Haenlein M., Kaplan A. (2019), A brief history of artificial intelligence: On the past, present, and future of artificial intelligence, *California Management Review*, 61, n.4, pp.5-14

Hobsbawm E.J. (1954), The General Crisis of the European Economy in the 17th Century, *Past & Present*, 5, n.1, pp.33-53

Holmes N.G., Wieman C.E., Bonn D.A. (2015), Teaching Critical Thinking, *Proceedings of the National Academy of Sciences*, 112, n.36, pp.11199-11204

Hong S.H. (2015), Subjunctive and interpassive "knowing" in the surveillance society, *Media and Communication*, 3, n.2, pp.63-76

Isceri M., Luppi R. (2022), L'impatto dell'intelligenza artificiale nella sostituzione dei lavoratori: riflessioni a margine di una ricerca, *Lavoro Diritti Europa*, 1, pp.2-17

Isceri M., Macioce F. (2022), Innovazione tecnologica e fiducia nei luoghi di lavoro. Problemi e prospettive giuridiche, in Robasto D. (a cura di), *Robot e cobot nell'impresa e nella scuola*, Milano, Franco Angeli, pp.25-34

Lee J.D., Moray N. (1992), Trust, control strategies and allocation of function in human-machine systems, *Ergonomics*, 35, pp.1243-1270

Lee K., Brownstein J.S., Mills R.G., Kohane I.S. (2010), Does Collocation Inform the Impact of Collaboration?, *PLoS ONE*, 5, pp.1-6

Plesner U., Husted E. (2022), *L'organizzazione digitale*, Bologna, il Mulino

Raisch S., Krakowski S. (2021), Artificial intelligence and management: The automation–augmentation paradox, *Academy of Management Review*, 46, n.1, pp.192-210

Salento A. (2018), Industria 4.0 e determinismo tecnologico, in Salento A. (a cura di), *Industria 4.0. Oltre il determinismo tecnologico*, Bologna, TAO Digital Library, pp.6-22

Schwab K. (2016), *La quarta rivoluzione industriale*, Milano, Franco Angeli

Tamburrini G. (2020), *Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiale*, Torino, Carocci

Wallace J. (2021), Getting collaborative robots to work: A study of ethics emerging during the implementation of cobots, *Paladyn. Journal of Behavioral Robotics*, n.12, pp.299-309

Wiener N. (1966), *Introduzione alla cibernetica. L'uso umano degli esseri umani*, Torino, Boringhieri

World Economic Forum (2019), *The Future of Jobs Report*, Ginevra, World Economic Forum

Zambonelli F. (2020), *Algocrazia. Il governo degli algoritmi e dell'intelligenza artificiale*, Trieste, Scienza Express Edizioni

Sara Pautasso
sarapautasso@hotmail.it
Assegnista di ricerca presso il Dipartimento di Giurisprudenza dell'Università di Pisa, dove ricopre il ruolo di co-docente del corso di Filosofia del diritto. I suoi studi si concentrano sulle implicazioni etiche e giuridiche dell'intelligenza artificiale. Fra le pubblicazioni recenti si segnalano: Giustizia, responsabilità, singolarità. Riflessi etico-giuridici dell'intelligenza artificiale a partire da Emmanuel Levinas, *Filosofia morale*, 2025, e *Tempi autoreferenziali. Una teoria critica del tempo moderno a partire dal pensiero di Emmanuel Levinas*, 2024.

Intelligenza artificiale come policy capacity
Sfide e promesse per la Pubblica amministrazione italiana

Fortunato Musella
Università degli Studi di Napoli Federico II

Luigi Rullo
Università degli Studi di Napoli Federico II

L'Intelligenza artificiale sta trasformando la Pubblica amministrazione con applicazioni in numerose aree, tra cui sanità, sicurezza e giustizia. Tuttavia, le politiche digitali faticano a tenere il passo dell'innovazione guidata dal settore privato. Questo articolo esamina alcune esperienze di IA nel settore pubblico italiano, evidenziando il loro impatto sul ciclo delle politiche pubbliche. Il concetto di *policy capacity* aiuterà a comprendere le sfide e le promesse dei percorsi di innovazione in corso, mostrando come 'amministrare con l'IA' richieda un cambiamento di paradigma nel set di competenze e risorse della PA.

Artificial intelligence is transforming Public administration, with applications across multiple sectors including healthcare, security, and justice. However, digital policies struggle to keep pace with innovation driven by the private sector. This article examines a series of AI implementations in the Italian public sector, highlighting their impact on the policy cycle. The concept of policy capacity helps to understand the challenges and opportunities of current innovation pathways, showing how 'administering with AI' requires a paradigm shift in the set of skills and resources of Public administration.

DOI: 10.53223/Sinappsi_2025-02-9

Citazione
Musella F., Rullo L. (2025), Intelligenza artificiale come policy capacity. Sfide e promesse per la Pubblica amministrazione italiana, *Sinappsi*, XV, n.2, pp.111-127

Parole chiave
Intelligenza artificiale
Policy capacity
Pubblica amministrazione

Keywords
Artificial intelligence
Policy capacity
Public administration

Introduzione

Nell'universo digicratico in evoluzione (Calise e Musella 2023), l'affermarsi dell'Intelligenza artificiale (IA) segna l'approdo in una nuova era e una sfida cruciale per le amministrazioni pubbliche come le abbiamo conosciute finora (Wirtz *et al.* 2022; Musella 2021; Meijer *et al.* 2021; Chandra e Feng 2025; Albrecht 2025). Innovazioni come l'analisi predittiva, l'IA generativa, o il riconoscimento facciale segnano ulteriori solchi in un precorso di algoritmizzazione di un ampio spettro di funzioni amministrative, suscitando reazioni contrastanti, tra preoccupazioni legate alla violazione della privacy, all'*explainability*, all'interoperabilità, alla cybersicurezza (Engstrom e Ho 2021), ed entusiasmi per le possibilità offerte, dal processo decisionale all'efficientamento dei servizi erogati (Criado *et al.* 2025). Esperienze recenti hanno dimostrato il potenziale trasformativo dell'impiego di IA nel settore pubblico europeo, con ambiti di applicazione che spaziano dalla sicurezza pubblica alla prevenzione delle catastrofi, dalla gestione del traffico ai servizi sanitari, fino alla regolamentazione dei

processi giudiziari e all'assegnazione di incarichi nelle scuole. Tuttavia, pensare che l'introduzione dell'IA basti, di per sé, a rafforzare il settore pubblico sarebbe fuorviante. Non produce, infatti, effetti omogenei o automatici, ma si configura come un fattore che ristruttura gradualmente relazioni di potere e flussi informativi, sulla base del contesto istituzionale e regolatorio, delle interdipendenze organizzative e della disponibilità di risorse per governarla (Di Giulio e Vecchi 2023). L'ascesa della governance algoritmica, dunque, impone un confronto più maturo sulla ristrutturazione dei presupposti funzionali, i valori fondanti e gli strumenti operativi dell'amministrazione pubblica (Musella 2024).

Il cantiere della riforma digitale ha ormai storia pluridecennale, che ha visto il maturare di rilevanti riflessioni e di prime implementazioni anche nel nostro Paese (Natalini 2022; Di Mascio *et al.* 2025). A partire dagli anni Novanta, lo sviluppo dell'*e-government* aveva impresso una svolta significativa al rinnovamento della Pubblica amministrazione, elevando l'uso delle ICT a elemento cardine del *reinventing government* (Osborne e Gaebler 1992). Negli anni Duemila il mantra *efficienza, efficacia, economicità* si affianca, prima, a quelli di pubblicità e trasparenza e, poi, a quello dell'accountability, ridisegnando l'intero circuito burocratico sulla base dei nuovi parametri resi possibili dalla rapidità e diffusività di una Pubblica amministrazione capace di viaggiare alla velocità della luce (Calise e Musella 2019, 89). La digitalizzazione, ulteriormente sostenuta dall'alterazione per qualità e quantità dei dati a disposizione, si presenta spina dorsale del rinnovamento del settore pubblico (Dunleavy e Margetts 2023), tanto da condurre alla diffusione di una concezione più ampia di governo elettronico in grado di abbracciare sia la modernizzazione amministrativa, sia il rinnovamento delle forme della democrazia (Amoretti e Musella 2012; per una panoramica si veda Di Mascio e Natalini 2022). Innovazioni che hanno assunto ancora maggior rilievo alla luce della crisi pandemica che ha imposto gravi interruzioni al funzionamento della Pubblica amministrazione e accelerato un suo ripensamento in chiave digitale, ma con risultati che molto spesso non hanno risposto alle aspettative tanto che "l'ottimismo sulla capacità taumaturgica della digitalizzazione anche fuori dall'Italia si è dimostrato, alla luce dei primi studi empirici, largamente ingiustificato" (Torchia 2021).

Se le sperimentazioni digitali incrementavano in numero ed estensione negli ultimi anni, è con l'IA che avviene il salto di scala. La consapevolezza dell'impatto che il digitale ha avuto in numerosissime sfere di attività pubblica e privata diventa allora un vero e proprio cambio di paradigma, con numerose applicazioni, come si vede nella tabella 1, a intervenire nelle diverse aree di policy. Con connessi rischi esistenziali per alcuni attori e istituzioni della nostra contemporaneità, sfidati dal veloce e inarrestabile sviluppo delle tecnologie avanzate.

In questo scenario, fronteggiare – e sfruttare – le opportunità più recenti della trasformazione digitale invita all'adozione di lenti analitiche che permettano di comprendere adeguatamente i mutamenti in corso, al confronto non solo con le sfide etiche (Smuha 2025) e regolative (Radu 2021) connesse all'IA, quanto con le sue potenzialità applicative. Resta poi il dato delle difficoltà relative all'implementazione di tecnologie che sono state adottate in ambito privato, con ritmi di diffusione che hanno reso le Corporation del tech attori globali di amplissimo potere economico e politico e che, invece, in ambito pubblico uniscono a eccellenze ed esperienze di successo un lento adeguamento delle pratiche e delle competenze.

In questo articolo leggeremo tali processi alla luce del concetto di *policy capacity* (Wu *et al.* 2015; 2018), definito come l'insieme di competenze, abilità e risorse delle amministrazioni per progettare e perseguire obiettivi di policy. Questo concetto offrirà una lente utile per interpretare criticamente l'impatto delle nuove tecnologie sulla Pubblica amministrazione, ricordandoci della loro ambivalenza intrinseca. Spesso, infatti, l'IA è stata narrata come una soluzione per razionalizzare la Pubblica amministrazione, tagliare il personale e automatizzare le decisioni, affidandole a calcoli algoritmici che escludono il giudizio umano. Questa impostazione, tuttavia, vede la macchina come superiore o sostitutiva dell'uomo, riducendo la complessità dell'azione pubblica a un problema computazionale (Sgueo 2024). Di conseguenza, è necessario un ribaltamento di prospettiva: invece di interpretare l'IA come un mero strumento per 'fare di più con meno', si suggerisce che il suo uso possa essere valorizzato solo in un contesto di *human-AI*

Tabella 1. Area di policy delle applicazioni IA nel settore pubblico in Europa (ricodifica COFOG livello 2)

Policy area	N.	%
Servizi generali delle pubbliche amministrazioni	482	28,9%
Affari economici	309	18,5%
Ordine pubblico e sicurezza	242	14,5%
Sanità	204	12,2%
Protezione sociale	127	7,6%
Protezione ambientale	105	6,3%
Istruzione	86	5,2%
Abitazioni e assetto del territorio	57	3,4%
Attività ricreative, culturali e di culto	47	2,8%
Difesa	9	0,5%
Totale	1.668	100%

Fonte: elaborazione degli Autori su dati forniti da European Commission, Joint Research Centre (JRC) (2023)

interaction, elevandola a strumento di innovazione istituzionale (Musella 2021). Evitando una visione deterministà dell'innovazione tecnologica nel settore pubblico, il concetto di *policy capacity* permetterà da un lato di considerare cosa manca alla Pubblica amministrazione, in particolare italiana, per una più celere ed efficace implementazione delle politiche digitali basate sull'IA, dall'altra di evidenziare come è la stessa IA a potersi leggere come uno strumento e un moltiplicatore di *policy capacity*. In breve, un'occasione, forse non ripetibile, per l'amministrazione pubblica per essere al passo con il vortice di cambiamento in corso.

1. La metamorfosi digitale della PA

L'IA sta progressivamente ridefinendo il *public management* grazie a numerose iniziative che testimoniano – su scala globale – il crescente impatto di questa tecnologia nella gestione della cosa pubblica. Sebbene il processo muova ancora i suoi primi passi (Pencheva *et al.* 2020), la possibilità di 'amministrare con l'IA' ha reso più ambiziosa la trasformazione digitale della Pubblica amministrazione, che è divenuta una priorità delle strategie di sviluppo nazionali e sovranazionali[1], in particolare in tema di accesso e gestione dei dati e di principi etici nella sua progettazione (van Noordt *et al*. 2025, 226-229). Non si tratta più solo di rendere i servizi più efficaci ed efficienti, ma di mettere le potenzialità dell'IA a disposizione della funzione pubblica e fare i conti con la necessità di un adeguamento radicale delle amministrazioni nel nuovo scenario (Howlett *et al.* 2025; Valle-Cruz *et al.* 2020).

La capacità dell'IA di generare, analizzare e sintetizzare informazioni ha già aperto nuove vie alla *public innovation*. Strumenti come Archibot semplificano la consultazione degli archivi del Parlamento europeo[2], applicazioni come BERT aiutano ad analizzare testi normativi, facilitando la comprensione e il confronto delle leggi tra gli Stati membri dell'Unione europea attraverso l'analisi avanzata del linguaggio naturale. Nel Regno Unito AI tool come Redbox stanno supportando i funzionari pubblici del *Cabinet Office* e del *Department for Science, Innovation and Technology* a riassumere

1 Basti considerare, in Europa, il recente lancio di InvestAI, un'iniziativa volta a mobilitare 200 miliardi di euro per investimenti in IA, compreso un nuovo fondo europeo di 20 miliardi per la gigafactory di IA. Si noti che, ad oggi, nessuna tra le *big giant* globali ha 'cittadinanza europea', alimentando così il *platform gap* da Stati Uniti e Cina. Dati, strumenti e approfondimenti utili a mappare in maniera costante questo terreno d'analisi sono forniti dall'OECD AI Policy Observatory (https://oecd.ai/en/dashboards/overview). Inoltre, si veda il *G7 toolkit for Artificial intelligence in the public sector*, https://assets.innovazione.gov.it/1728922597-g7-toolkit-for-ai-in-the-public-sector.pdf.

2 Sfruttando Claude di Anthropic, società americana supportata da Google e Amazon, il Parlamento europeo ha introdotto *Ask the EP Archives* (Archibot), un assistente IA che ha reso più accessibile la consultazione degli archivi. La stessa Anthropic, inoltre, ha firmato nel febbraio 2025 un memorandum d'intesa con il governo inglese guidato da Keir Starmer per introdurre servizi IA specializzati per l'amministrazione pubblica.

rapidamente le raccomandazioni di policy dai documenti ufficiali, mentre Parlex sta rendendo più efficiente la preparazione di disegni di legge[3] (Urban 2025; Vogl *et al.* 2020). Fuori dei confini europei, a Singapore, Pair Chat fornisce assistenza nella scrittura di e-mail, nella ricerca di informazioni e nella gestione documentale per i funzionari governativi grazie a una chatbot IA[4]. Negli Stati Uniti, il Governo di Donald Trump sta rimodellando la General Services Administration (GSA), con l'obiettivo di trasformarla in una startup del software. Questa metamorfosi, che si inserisce nel quadro più ampio di ristrutturazione dell'amministrazione americana, prevede l'automazione di numerosi processi interni e la centralizzazione dei dati federali, con uno dei primi pilastri rappresentato da una chatbot di IA generativa (GSAi) progettata per ridefinire e potenziare il lavoro (Dave *et al.* 2025). La Cina, inoltre, sta sondando le frontiere degli *AI civil servants* basate sulle architetture cognitive di *DeepSeek-R1*, così da impattare non solo le operazioni amministrative (es. elaborazione di documenti, smistamento dei compiti tra dipartimenti), ma anche il ridimensionamento degli impiegati del settore. Il progetto si pone in continuità con l'obiettivo di instaurare una profonda sinergia nelle relazioni tra governo e aziende private con il primo che "mira sempre più a diventare un creatore di piattaforme, un facilitatore e un promotore di strategie per l'innovazione, piuttosto che un attore o uno sponsor principale" (Todaro 2024, 78).

Guardando al caso italiano, un recente lavoro di monitoraggio condotto dalla Commissione europea evidenzia come Ministeri e agenzie pubbliche stiano introducendo sperimentazioni in grado di automatizzare compiti ripetitivi, migliorare la previsione e il monitoraggio dei rischi, personalizzare i servizi per i cittadini e ottimizzare l'allocazione e la gestione delle risorse (tabella 2). Un'ambizione poi ulteriormente sostenuta dagli investimenti finanziari e progettuali del PNRR e dagli obiettivi delineati nella *Strategia italiana per l'Intelligenza artificiale* (Agid 2024), che punta a implementare 400 progetti di IA entro il 2026[5] e, al tempo stesso, ovviare a legacy di lungo periodo che ne hanno rallentato lo sviluppo (Musella 2022; Natalini 2022).

In questo scenario, pur riconoscendo la natura spesso frammentaria di un'area di studio al crocevia tra diverse discipline e specializzazioni (scienza politica, management, diritto ecc.), si possono evidenziare almeno tre importanti direttrici delle politiche digitali, che sono anche le tre principali direzioni del loro sviluppo: integrazione, personalizzazione e regolazione.

In primo luogo, la letteratura si è andata arricchendo di osservazioni relative allo sviluppo dell'IA e alla sua integrazione nei processi amministrativi per incrementarne l'efficienza. Dinamiche che hanno raggiunto una massificazione e trasversalità del suo impiego, marcando una reingegnerizzazione dei flussi di lavoro grazie al proliferare di nuovi tool che permettono sia il miglioramento dell'accesso e la gestione avanzata di grandi quantità di dati, che la loro elaborazione, così da garantire una progressiva semplificazione del processo decisionale. Accelerando, ad esempio, valutazioni per determinare l'idoneità all'accesso a politiche redistributive, velocizzare le pratiche di richiesta d'asilo, o localizzare interventi edilizi non autorizzati (come avvenuto nel caso delle piscine abusive in Francia grazie a un algoritmo progettato da Google). Se da un lato i benefici attesi della

3 Queste innovazioni suscitano crescenti entusiasmi per migliorare la qualità troppo spesso deficitaria del *lawmaking*. In quest'ottica, si vedano i casi di Ulysses in Brasile, QuantGov negli Stati Uniti, LLaMandement in Francia. In Italia si veda il progetto pubblicato dal Comitato di vigilanza sull'attività di documentazione (2024) *Utilizzare l'intelligenza artificiale a supporto del lavoro parlamentare* e il lancio, nell'estate del 2025, di prototipi di IA generativa per i lavori della Camera dei Deputati. Inoltre, si veda il progetto Savia, un programma che nasce dalla collaborazione tra la regione Emilia-Romagna e Cineca, che mira a sperimentare prototipi a "disposizione di chi fa le leggi per valutarne in anticipo impatto ed efficacia, a partire dalla consultazione semplice e veloce delle banche dati di leggi e atti amministrativi regionali". I benefici della GenAI nell'attività di *lawmaking* restano, tuttavia, condizionati dalla presenza di bias e allucinazioni, che non sono semplici malfunzionamenti tecnici, ma il prodotto di tendenze cognitive e culturali incorporate nei dati di addestramento (Cantens 2025). Queste distorsioni, una volta trasposte in strumenti generativi nell'attività di *lawmaking*, rischiano di essere amplificate in modo sistematico, diffondendo contenuti potenzialmente incompatibili con i diritti fondamentali (Rangone e Megale 2025).

4 I dati ufficiali (https://pair.gov.sg/products/chat) evidenziano che, con oltre 96.000 utenti in 207 agenzie pubbliche, l'introduzione di Pair Chat ha ridotto i tempi di lavoro del 49%, gestendo 29,4 milioni di interazioni.

5 Il *Piano Triennale per l'informatica nella Pubblica amministrazione 2024-2026* (Agid 2023) presenta un utile approfondimento su questi temi. Si veda, inoltre, il Rapporto Agid (2025).

Tabella 2. Processi che utilizzano l'IA nel settore pubblico in Europa e in Italia[6]

	Europa		Italia	
	N.	%	N.	%
Aggiudicazione	37	2,2%	3	1,5%
Analisi, monitoraggio e ricerca normativa	417	25,0%	39	19,8%
Enforcement	348	20,9%	33	16,8%
Management interno	266	15,9%	30	15,2%
Servizi al pubblico e comunicazioni	600	36,0%	92	46,7%
Totale	1.668	100,0%	197	100,00%

Fonte: elaborazione degli Autori su dati forniti da European Commission, Joint Research Centre (JRC) (2023)

algorithmic regulation non sempre si concretizzano nella pratica con evidenti rischi di compressione dei diritti individuali[7] (es. SyRi, Compas, Buona Scuola, Gladsaxe ecc.) (Smuha 2025), l'ottimizzazione dell'efficacia ed efficienza dei servizi pubblici può avere ricadute positive sia sulla capacità delle istituzioni di analizzare una mole di dati sempre più abbondante che sulla produttività, grazie a un progressivo snellimento delle operazioni del personale della Pubblica amministrazione (Misuraca e van Noordt 2020).

In secondo luogo, l'IA migliora la personalizzazione dei servizi pubblici utilizzando i dati per adeguarli alle preferenze dei cittadini, così da favorire risposte calibrate sulle specificità dei bisogni (van Ooijen *et al.* 2019). Al tempo stesso, a mano a mano che i dati a disposizione delle amministrazioni aumentano in dettaglio e precisione sta diventando sempre più agevole identificare caratteristiche socio-demografiche come età, genere, residenza e valutare l'impatto delle politiche pubbliche sui diversi gruppi. Ma anche aggredire l'estrema varietà di questi dati digitali che restituiscono comportamento e preferenze dei cittadini, a partire dalle tracce che essi lasciano sul web e che permettono forme avanzate di *micro-targeting*. Si tratta di uno dei fulcri dell'innovazione nel settore tecnologico privato (The Economist 2024) e che può guidare, con mutate logiche orientate alla creazione di valore pubblico e in linea con i diritti fondamentali, anche i servizi pubblici offerti, affinché i cittadini abbiano la sensazione che lo Stato lavori davvero per loro (Margetts *et al.* 2024, 58; Tangi *et al.* 2024) e incrementino la loro fiducia nelle istituzioni (Haesevoets 2024). Si pensi, in particolare, al moltiplicarsi delle chatbot e degli assistenti virtuali che, da un lato, facilitano una maggiore attenzione dei burocrati nella cura di servizi più complessi, a pieno vantaggio dei cittadini (Salah *et al.* 2023). Dall'altro, accanto alla maggiore flessibilità di una PA aperta 24/24, l'impiego delle chatbot potrebbe generare nei cittadini sentiment negativi legati al rischio che l'interazione con l'IA sia più rigida dell'azione umana, non riuscendo a prevedere quelle soluzioni di compromesso spesso necessarie nell'adattamento di regole e criteri generali ai casi specifici (Cortés-Cediel *et al.* 2023).

In terzo luogo, l'adozione dell'IA sta alimentando il dibattito pubblico e scientifico non solo sulla necessità di monitorare – e indirizzare – i mutamenti in atto ma anche con l'urgenza di regolarne pubblicamente l'utilizzo, in linea con l'ideale democratico per cui senza trasparenza non può esserci responsabilità (Bobbio 1980). Basti considerare come il mercato privato delle soluzioni di AI policy advisory sia in rapida espansione, trainato da Corporation come Google, Open AI, Amazon, McKinsey, IBM. Strumenti avanzati come Google AI for Public Sector, Amazon Lex, IBM Watson

6 Più nel dettaglio, i processi amministrativi che utilizzano l'IA includono l'aggiudicazione di benefici o diritti, l'analisi e monitoraggio delle politiche pubbliche, l'*enforcement*, anche predittivo, delle norme. La gestione interna supporta il funzionamento dell'ente, mentre i servizi al pubblico riguardano le misure finalizzate a semplificare l'accesso, personalizzare l'esperienza e rafforzare il coinvolgimento e la comunicazione con i cittadini.

7 L'OCSE ha recentemente lanciato l'AI Incidents Monitor (https://oecd.ai/en/incidents), un database che traccia i principali fallimenti, abusi e incidenti legati all'IA in tutto il mondo. Questa piattaforma ha lo scopo di documentare i danni reali causati dall'IA in diversi ambiti, tra cui violazione della privacy, discriminazione, disinformazione e problemi di sicurezza (Santoriello 2025).

Government AI, stanno plasmando il modo in cui i dati e le informazioni vengono analizzate, riducendo il tempo necessario per raccogliere e sintetizzare informazioni, spingendosi a valutare la coerenza con gli impianti normativi delle decisioni adottate. Ad esempio, su ChatGPT sono presenti servizi espressamente dedicati ai policy maker come FOIAbot o Policy advisor che assicurano funzioni di policy analysis, strategic advice, identificazione di opportunità di innovazione, interpretazione dei dati, ricerca di policy di successo in diverse regioni e settori potenzialmente (ri)adattabili sulla base delle proprie esigenze. A ciò si aggiunge il pericolo di un'eccessiva espansione degli attori privati che potrebbero esercitare un controllo indiretto ma significativo sulle infrastrutture decisionali pubbliche, imponendo standard, modelli predittivi e logiche di ottimizzazione non sempre coerenti con i valori democratici e/o rendendo difficilmente ispezionabili gli stessi sistemi. In questo contesto, la discussione tra leader del settore, policy maker, accademici e rappresentanti della società civile non guarda più a quando le amministrazioni adotteranno queste soluzioni, ma su come e con quali tutele per i cittadini riusciranno a farlo. Dietro l'immagine di un'amministrazione superpotente, infatti, sono sempre più evidenti i rischi – di natura sociale, etica, normativa, tecnologica (Wirtz e Müller 2019) – che possono minare la trasparenza amministrativa, "con il risultato della traslazione di potere decisionale dalla mano pubblica a quella privata, verso i nuovi dominus in grado di sostituire al codice della legge quello dell'algoritmo" (Musella 2021, 105; Khanal *et al.* 2025). D'altro canto, il non adeguamento della Pubblica amministrazione al corrente stato di avanzamento tecnologico che sta impattando sulla società porterebbe a una chiara svalutazione del pubblico, sulla base della constatazione di una sua palese incapacità di azione.

È necessario, dunque, il bilanciamento tra la capacità di costruire un'infrastruttura regolatoria che non solo limiti i rischi, ma sappia valorizzare le opportunità trasformative dell'IA in chiave democratica. Tra i diversi (macro)approcci in campo, l'Unione europea ha posto al centro dell'agenda che la grande trasformazione in cui siamo immersi non può essere solo tecnologica, ma anche umanistica ed etica, offrendo strumenti che spaziano dall'introduzione di standard qualitativi alla necessità di produrre documentazione tecnica specifica per dimostrare la responsabilità legale e la conformità alle normative esistenti (Gdpr, Digital services Act, Digital market Act, Data governance Act, Artificial intelligence Act - AI Act). In particolare, grazie all'AI Act, l'obiettivo è quello democratizzare e armonizzare l'AI, limitarne i rischi e garantire che i modelli di IA siano giustiziabili e improntati ai diritti umani e alla trasparenza, attraverso "un quadro giuridico unico per lo sviluppo dell'Intelligenza artificiale di cui ci si può fidare"[8].

Il Regolamento europeo si fonda infatti su una logica di regolazione basata sul rischio, distinguendo tra diversi livelli di criticità (inaccettabile, alto, limitato, minimo) e imponendo obblighi proporzionati alla pervasività e al potenziale impatto dei sistemi d'IA. In generale, l'Europa intende limitare un ampio spettro di rischi che vanno dalla discriminazione algoritmica all'opacità decisionale, dall'erosione della privacy alla compromissione dei diritti individuali. In particolare, è vietato l'uso intrusivo e discriminatorio di applicazioni rivolte all'identificazione biometrica da remoto, 'in tempo reale' e 'a posteriori', come i sistemi di categorizzazione biometrica basati su caratteristiche sensibili (es. genere, etnia ecc.), i sistemi di polizia predittiva e quelli di riconoscimento facciale. L'AI Act, dunque, affronta questi scenari riconoscendo che l'IA non è strumento neutrale, ma incorporata in architetture di potere che possono rafforzare disuguaglianze esistenti o generare nuove forme di esclusione (Santoriello 2025). Riflette, quindi, l'esigenza di trasparenza e supervisione pubblica, riaffermando l'idea che il controllo su tecnologie ad alto impatto non può essere lasciato esclusivamente alle Corporation. Più che una normativa tecnica si configura, dunque, come una cornice politica per l'articolazione di uno sviluppo dell'IA compatibile con i diritti fondamentali e con una visione etica della sua progettazione.

Tuttavia, le ambizioni dell'AI Act si scontrano con una serie di criticità strutturali e operative che mettono in risalto lo scarto tra la sfera dei desiderata e l'effettiva capacità di regolazione. In primo luogo, vi è il rischio che l'impianto normativo – pur tra i più

8 Queste le parole della Presidente della Commissione europea Ursula Von der Leyen in un post su X dell'8 dicembre 2023, https://twitter.com/vonderleyen/status/1733257654254883095.

avanzati al mondo – non sia sufficiente a contenere le dinamiche di concentrazione del potere nelle mani di pochi attori privati, capaci di influenzare standard tecnici, modelli di business e agende politiche su scala transnazionale. In secondo luogo, l'efficacia della regolazione dipende dalla concreta capacità dei governi di applicare i nuclei del regolamento che, tuttavia, "appare particolarmente variabile e soggetta a fattori storici e politici: ne consegue che la concezione di rischio legata ad un dato sistema di Intelligenza artificiale potrebbe mutare nel tempo" (Armiento 2025, 25). Ancor di più in un contesto segnato da profonde asimmetrie di competenze, risorse e conoscenze tecniche tra potere privato e potere pubblico (Ferrarese 2022; Sgueo 2024). Come sostenuto dallo stesso Parlamento europeo "data la natura dei modelli di base, mancano competenze specifiche in materia di valutazione della conformità e i metodi di audit di terze parti sono ancora in fase di sviluppo"[9]. A ciò si aggiunge un elemento fondamentale, sempre più discusso dalla letteratura più recente (van Dijck *et al.* 2025): la regolazione, per quanto dettagliata, non può da sola contenere la complessità socio-politica dell'IA.

Da questo punto di vista, l'IA non è soltanto un oggetto da regolare, ma diventa anche un moltiplicatore di *policy capacity*, a seconda di come viene adottata, con quali finalità e con quali competenze a supporto.

2. Una questione di policy capacity

L'integrazione delle tecnologie di IA sta permettendo ad alcune amministrazioni di creare ambienti favorevoli alla sperimentazione e progettare applicazioni che stanno raggiungendo risultati che intersecano – e riscrivono – con modalità e forza differenti il modo in cui gli amministratori pubblici progettano, implementano e valutano le politiche pubbliche.

Il framework teorico della *policy capacity* formalizzato da Wu *et al.* (2015; 2018) aiuta sia a comprendere l'auspicabilità che l'adozione dell'IA nella Pubblica amministrazione italiana comporta, sia come la sfida ad essa legata non è solo relativa agli aspetti di natura tecnica, ma anche alla riscrittura della caratterizzazione organizzativa e culturale della Pubblica amministrazione. Gli autori, infatti, sostengono che la *policy capacity* descrive "l'insieme di competenze e risorse – o competenze e capacità – necessarie per svolgere *policy function*" (Wu *et al.* 2018, 3) e presentano una griglia analitica che distingue nella *policy capacity* di un'amministrazione tre skill critiche essenziali (analitica, operativa, politica) e tre risorse critiche lungo tre livelli (individuale, organizzativo, sistemico), la cui intersezione dà vita a nove combinazioni che aiutano a leggere e interpretare il successo – o meno – di una politica pubblica (tabella 3).

In primo luogo, per capacità analitica si intende quella di "garantire che le azioni di policy siano tecnicamente valide, nel senso che possono contribuire al raggiungimento degli obiettivi politici se realizzate" (Wu *et al.* 2015, 167-168; Woo 2021, 33). In secondo luogo, la capacità operativa insiste sull'allineamento delle risorse "alle azioni di policy in modo che possano essere attuate nella pratica" (Wu *et al.* 2015, 168), così da poter tradurre le risorse disponibili in politiche attuabili (Woo 2021, 34). In terzo luogo, la capacità politica guarda alla capacità di "ottenere il sostegno politico per le azioni di policy" (Wu *et al.* 2015, 168), rivelandosi necessaria per garantire che le politiche abbiano l'approvazione e la legittimità necessarie per il successo (Woo 2021, 36).

Il punto di forza di questo framework è la sua capacità di coprire l'intero ciclo di policy. Pur essendo strettamente interconnesse, la *policy capacity* nelle dimensioni analitiche, operative e politiche risponde a logiche diverse, offrendo contributi distinti ma indispensabili al processo decisionale (Mukherjee *et al.* 2021). Leggere, quindi, questi processi attraverso questa prospettiva permette non solo di individuare con maggiore precisione i fattori di successo e le criticità, ma anche di facilitare l'adattamento e il miglioramento continuo delle strategie per l'integrazione dell'IA nel settore pubblico (Kim *et al.* 2023; Khan e Hussain 2024), così da comprendere cosa la PA dovrebbe fare o essere in grado di fare.

Alla luce di queste premesse, l'introduzione dell'IA nella Pubblica amministrazione rappresenta una risorsa complessa a tutti i livelli e richiede, soprattutto in questa fase, un forte impegno a

9 Si veda il comunicato stampa del 14 giugno 2023, *MEPs ready to negotiate first-ever rules for safe and transparent AI*, https://www.europarl.europa.eu/news/en/press-room/20230609IPR96212/meps-ready-to-negotiate-first-ever-rules-for-safe-and-transparent-ai.

Tabella 3. Le dimensioni della Policy capacity

	Skill e competenze		
Livelli di risorse e capacità	*Analitico*	*Operativo*	*Politico*
Individuale	Capacità analitica individuale	Capacità operativa individuale	Capacità politica individuale
Organizzativo	Capacità analitica organizzativa	Capacità operativa organizzativa	Capacità politica organizzativa
Sistemico	Capacità analitica sistemica	Capacità operativa sistemica	Capacità politica sistemica

Fonte: adattato da Wu *et al.* (2015)

intraprendere un'attività di modellazione della *policy capacity* così da definire le direzioni strategiche, valutare le implicazioni delle alternative di policy e utilizzare in modo appropriato la conoscenza nel processo decisionale (Wu *et al.* 2015; 2018; Misuraca *et al.* 2024). Diventa, dunque, uno strumento chiave per elaborare policy in modo più efficace ed efficiente e incrementare innanzitutto le capacità analitiche dei singoli e delle organizzazioni nel conseguire policy di successo nel breve e medio periodo (Capano *et al.* 2025). Si tratta di un apporto fondamentale considerato che, come ricorda Howlett (2015, 174) "le organizzazioni con capacità analitiche più forti hanno quindi maggiori probabilità, *ceteris paribus*, di avere un impatto maggiore sui risultati rispetto a quelle che mancano delle componenti principali di tale capacità". Se a lungo i manager pubblici hanno operato con informazioni limitate, risorse scarne e/o con la possibilità di valutare solo un sottoinsieme di opzioni possibili, grazie all'IA la *policy capacity* è potenziata, sia in termini di velocità e quantità di dati, che per la torsione delle capacità individuali e organizzative in un'ottica *evidence-based*. Di fatto "le trasformazioni epistemiche introdotte dall'Intelligenza artificiale ridefiniscono tutti gli aspetti del lavoro politico, creando nuove razionalità, quadri di riferimento e schemi d'azione" (Filgueiras 2025, 54; Yar *et al.* 2024). La sua adozione, quindi, presenta un intreccio eterogeneo di ricognizione dei cambiamenti in atto e di adeguamento dei principi e delle procedure che guidano l'azione pubblica e, con diversi livelli di rischio, la possibilità di interpretare e utilizzare dati.

In primo luogo, gli amministratori pubblici sono i principali attori che hanno a disposizione nuovi mezzi per identificare *issue* e ascoltare le esigenze della cittadinanza, così da riorganizzare i loro interventi di policy sulla base di una maggiore capacità di raccolta e analisi delle informazioni pubbliche[10]. Sfruttando l'IA e, in particolare, tecniche di *machine learning*, istituti e agenzie pubbliche possono tracciare, quasi istantaneamente, argomenti e tendenze emergenti e definire i punti all'ordine del giorno. Un esempio concreto è il progetto che ha visto coinvolto l'Istituto sanitario nazionale, che ha sviluppato nel 2020 nell'ambito del progetto europeo Eu-Jav, una piattaforma per fronteggiare la pandemia da Covid-19. Basandosi su dati raccolti dalla Rete (es. Twitter (oggi X), Reddit, Wikipedia, Google Trends), il progetto rientrerebbe tra i sistemi a rischio minimo dell'IA act perché non impatta direttamente sui diritti individuali, permettendo di definire – e localizzare – in maniera puntuale le strategie di comunicazione per la promozione delle vaccinazioni (Rizzo *et al.* 2021). Strumenti che aiutano a individuare problemi emergenti, segnalarne l'urgenza e stabilire priorità di azione sono stati al centro del progetto europeo Ethical technology adoption in Public administration Services (ETAPAS), un'iniziativa che ha coinvolto il MEF con l'obiettivo di identificare i rischi etici e di privacy legati ai dati aperti della piattaforma NoiPA e fornire specifiche raccomandazioni e azioni di mitigazione da mettere al centro dell'agenda.

In secondo luogo, l'introduzione dell'IA e, in particolare, di modelli di IA generativa, consente di potenziare la *policy capacity* analitica da parte delle amministrazioni, migliorando l'efficacia nelle risposte attraverso la raccolta in tempo reale di dati strutturati sulle criticità espresse dagli utenti

10 L'Unione europea, ad esempio, ha adottato DORIS (Data, opinions and reports information system) che analizza e sintetizza i feedback ricevuti durante le consultazioni pubbliche. Il tool utilizza il *natural language processing* per aggregare commenti e rilevare temi chiave così da supportare i lavori della Commissione.

(Mergel *et al.* 2024). È il caso, ad esempio, dell'Inps, uno degli istituti più proattivi sullo scenario italiano[11] (Cacciatore 2024), che ha introdotto IA chatbot conversazionali come il 'Consulente digitale delle pensioni'. Tuttavia, la diffusione delle IA chatbot ai vari livelli dell'amministrazione (es. Urpy, Camilla per i concorsi pubblici gestiti del Formez, la chatbot del Ministero degli Esteri ecc.) evidenziano come questi sistemi possano sostanzialmente rappresentare – con rischio limitato – una fonte preziosa per analizzare i bisogni emergenti, identificare lacune informative e comprendere le dinamiche di accesso ai servizi pubblici, fornendo insight utili per orientare la formulazione di policy più aderenti alle reali esigenze dell'utenza. In questo senso, l'IA non si limita a ottimizzare processi di front office, ma contribuisce a costruire un sistema decisionale più adattivo e basato sulle esigenze dei cittadini. L'automazione delle interazioni, infatti, genera nuove basi informative che possono essere impiegate per progettare interventi più mirati, rafforzando così la componente analitica della *policy capacity* e aiutare l'ente ad anticipare scenari critici. Al tempo stesso, va sottolineato che questi strumenti devono garantire trasparenza verso l'utente, dichiarando esplicitamente che si sta interagendo con un sistema automatizzato e rendendo conoscibili i criteri alla base delle risposte suggerite. Inoltre, il loro utilizzo può accentuare forme di disuguaglianza, con alcune fasce della popolazione – come anziani o cittadini con bassa alfabetizzazione digitale – che rischiano di essere escluse o penalizzate, accentuando le disuguaglianze se non affiancate da canali alternativi (Larsen e Følstad 2024).

In terzo luogo, l'IA contribuisce a rafforzare la *policy capacity* analitica attraverso la produzione di *what-if scenario* e l'identificazione dei potenziali impatti delle misure adottate. Sogei sta testando diversi sistemi basati su IA e *natural language processing* (NLP), intensificando il suo utilizzo soprattutto nei processi di verifica della documentazione nelle richieste di subappalto. Sfruttando questi strumenti è possibile elaborare grandi quantità di documentazione (normative, regolamenti, richieste, documenti amministrativi), individuando pattern ricorrenti e problemi sistemici nel procurement pubblico, con il potenziale di anticipare tendenze e adattarvisi rapidamente. Di conseguenza, si rafforzano le capacità di fornire insight basati su dati strutturati, permettendo ai decisori pubblici di anticipare tendenze e adottare strategie di intervento più mirate. In maniera altrettanto innovativa, la Consob si è concentrata sull'efficientamento delle attività di monitoraggio e regolamentazione, introducendo algoritmi di lettura automatizzata dei documenti sintetici di informazioni (KIDs) utili per individuare eventuali abusi di mercato (Albè e Bottini 2022). Da segnalare, inoltre, l'esperienza dell'articolo 30 del Codice dei Contratti pubblici (art. 30, D.Lgs. 36/2023) che illustra l'avvio tangibile di un processo di ottimizzazione del processo decisionale, grazie a una transizione digitale dell'intero ciclo degli appalti pubblici. Spaziando, infatti, dalla programmazione alla gestione dell'esecuzione contrattuale, si introduce la possibilità di usare l'IA per rendere più efficienti le procedure di gara e gestire in maniera più trasparente l'aggiudicazione dei contratti pubblici (Licata 2024). Sul punto, tuttavia, resta centrale la necessità di "un contributo umano capace di controllare, validare, ovvero smentire la decisione automatizzata"[12], scongiurando il rischio che l'IA possa acquisire un ruolo sostanziale nella formazione della volontà procedimentale. In alcuni casi, come nella procedura di mobilità dei docenti prevista dalla riforma c.d. Buona Scuola, si era manifestata una criticità rilevante: risultava difficile stabilire se la decisione finale fosse semplicemente supportata dall'algoritmo o se coincidesse interamente con esso. Di conseguenza, come argomentato da diverse pronunce della giurisprudenza amministrativa (Dalfino 2020), l'azione dell'amministrazione rischiava di ridursi a una mera ratifica del risultato prodotto dal sistema automatizzato, compromettendo la trasparenza e la responsabilità del processo decisionale.

L'IA, inoltre, può offrire un contributo significativo in ambiti particolarmente delicati

11 Il documento *Linee guida sull'implementazione di sistemi di Intelligenza artificiale in Inps* presenta in dettaglio la strategia Inps (Direttiva del Direttore generale n. 8 dell'8 aprile 2024). Si pensi, in particolare, a come l'Istituto stia mirando a ottimizzare il flusso di PEC in arrivo attraverso l'adozione di un sistema automatizzato in grado di analizzare i contenuti di ciascuna PEC e indirizzare ogni richiesta all'operatore specializzato nella specifica materia. Questo strumento può essere riletto in ottica di *policy capacity* se interpretato come meccanismo per individuare criticità ricorrenti.

12 Per un'analisi approfondita del concetto anche in relazione all'art. 22 del GDPR, si veda Armiento (2025, 51-58).

come l'implementazione di politiche di prevenzione del crimine. Sperimentazioni come XLaw, Delia e Vigilium, sollevano interrogativi in termini di conformità rispetto al nuovo IA Act. Sarà centrale, quindi, verificare se esse rientrino tra le pratiche vietate – come la profilazione o il *social scoring* – oppure se possano essere adeguate ai requisiti previsti per i sistemi ad alto rischio, con particolare attenzione agli aspetti di trasparenza e verificabilità[13]. Questi strumenti, infatti, forniscono agli operatori di pubblica sicurezza informazioni predittive basate su dati aggregati provenienti da denunce, social network e fonti aperte, sollevando serie preoccupazioni in termini di protezione dei dati personali e rispetto della privacy (Armiento 2025, 69-81). XLaw, ad esempio, generando mappe di rischio su aree urbane, può produrre effetti distorsivi per cui l'intensificazione dei controlli in una determinata zona alimenta artificialmente la percezione di rischio, con potenziali conseguenze discriminatorie su specifici territori o gruppi sociali. Come sottolineato dall'esperienza americana nel caso Compas e più di recente nel caso del software KrimPro utlizzato dalla polizia di Berlino (Meijer *et al.* 2021), l'opacità algoritmica limita il diritto alla difesa, mina i principi del giusto processo, e rende problematico anche l'uso processuale degli output predittivi.

Altrettanto rischiose le sperimentazioni di tool IA nella lotta all'evasione fiscale. Si pensi al progetto Verifica dei rapporti finanziari-VeRa dall'Agenzia delle Entrate (Borghetti 2023), che introduce l'IA per aumentare la conformità fiscale e ridurre il livello di economia sommersa e il divario fiscale[14]. Basandosi su dati provenienti dall'incrocio di informazioni finanziarie e patrimoniali da diversi dataset pubblici (es. registri delle imprese, dati catastali), il progetto VeRa, da un lato, facilita l'attività di accertamento, dall'altro, mette in luce una serie di rischi per la gestione dei dati personali, in particolare in relazione al principio di minimizzazione dei dati previsto dal GDPR, trattando dati non sempre essenziali per le finalità fiscali. È fondamentale, dunque, legittimare adeguatamente l'uso dell'IA in ambiti ad alto rischio e, in particolare, evitare derive verso una 'presunzione algoritmica'. Quest'ultima potrebbe tradursi in segnalazioni ingiustificate e accertamenti errati o continui su determinate aree geografiche e/o categorie di contribuenti, rafforzando così pregiudizi sistemici e generando disuguaglianze. Inoltre, considerato che l'utilizzo dell'IA può introdurre *bias*, è essenziale che l'uomo sia sempre responsabile del procedimento e titolare della 'decisione finale' (Borghetti 2023, 106). Questo aspetto è stato recentemente richiamato nel DDL 1146 (*Disposizioni e delega al Governo in materia di Intelligenza artificiale*) in cui si ricorda all'art. 13 che l'utilizzo dell'IA nella Pubblica amministrazione è da intendersi "in funzione strumentale e di supporto all'attività provvedimentale, nel rispetto dell'autonomia e del potere decisionale della persona che resta l'unica responsabile dei provvedimenti e dei procedimenti in cui sia stata utilizzata l'Intelligenza artificiale".

In quarto luogo, l'IA sta accelerando una transizione verso forme di valutazione continua delle policy, aprendo la strada a rilevazioni – anche in tempo reale – di eventuali criticità, così da rafforzare l'azione dei manager pubblici nell'adozione di interventi correttivi (Pencheva *et al*. 2020). In quest'ottica, l'IA favorisce strutture di monitoraggio e automatizzazione delle valutazioni così come messo in luce dall'Inail, che la utilizza per l'analisi predittiva degli infortuni, il monitoraggio delle frodi e l'ottimizzazione dei servizi pubblici. Sistemi di *machine learning* analizzano dati su rischi, indennizzi e feedback degli utenti per misurare l'efficacia delle politiche e suggerire miglioramenti. In questo

13 Si pensi al caso del sistema SARI (Sistema automatico di riconoscimento immagini) che è attualmente impiegato dalla Polizia di Stato italiana per supportare le indagini attraverso il riconoscimento facciale attraverso *machine learning* basato su *deep learning*. In particolare, SARI Enterprise permette di confrontare un'immagine (frame o foto) con il database AFIS (che contiene le foto segnaletiche dei soggetti già identificati), ma a differenza della versione SARI Real-time (vietato dal Garante della Privacy nel 2021 in quanto considerato in contrasto con il GDPR) viene usato a posteriori su richiesta degli investigatori e non in tempo reale. L'esito restituisce una rosa di somiglianze che viene poi analizzata e validata manualmente da operatori specializzati.

14 Il progetto è stato lanciato nel 2022 e opera nel rispetto delle disposizioni della Legge di Bilancio del 2020 (articolo 1, commi 681-686, della legge 27 dicembre 2019, n. 160). Nel documento illustrativo della logica degli algoritmi, l'Agenzia delle entrate evidenzia come sistemi di IA "consentono di isolare rischi fiscali, anche non noti a priori, che, una volta individuati, possono essere utilizzati per l'elaborazione di autonomi criteri selettivi, ovvero permettono di attribuire una determinata probabilità di accadimento ad un rischio fiscale noto".

modo, l'IA non solo supporta la gestione operativa, ma fornisce dati utili per valutare l'efficacia delle policy esistenti e progettare interventi migliorativi, potenziando la *policy capacity* in chiave *evidence-based*. Tuttavia, considerata la natura sensibile dei dati trattati (sanitari e previdenziali), questi sistemi ricadono nella categoria ad alto rischio dell'IA Act, richiedendo conformità a requisiti stringenti in termini di (alta) qualità dei dati, *auditability*, trasparenza del processo decisionale e supervisione umana. Sul punto si consideri l'esperienza del tool Savio (un software di data mining utilizzato dall'Inps per individuare anomalie nei certificati medici e ottimizzare i controlli sui lavoratori) che mostra come l'IA, pur incrementando la *policy capacity* della PA, può scivolare in zone grigie per cui il successo di queste sperimentazioni non dipende solo dall'introduzione di queste innovazioni, ma anche dalla capacità di difendere e legittimare il loro utilizzo, soprattutto nelle aule giudiziarie. Nel 2018, infatti, il Garante per la protezione dei dati personali e il Tribunale di Roma lo avevano ritenuto uno strumento di profilazione individuale, richiedendo consenso e notifica. Tuttavia, la Cassazione (ord. 6177/2023) ha successivamente stabilito che Savio non creava profili permanenti, ma si limitava ad attribuire punteggi ai singoli certificati, escludendo quindi che si trattasse di una vera profilazione.

3. L'amministrazione Digital Twin

Veniamo ora a uno dei punti più avanzati delle applicazioni che stiamo considerando. I Digital Twin (DT) rappresentano una delle frontiere più ambiziose della *public innovation*[15]. Possono essere definiti fondamentalmente come "una rappresentazione virtuale di un sistema fisico (e dell'ambiente e dei processi ad esso associati) che viene aggiornata attraverso lo scambio di informazioni tra il sistema fisico e quello virtuale" (VanDerHorn e Mahadevan 2021, 2). Si tratta, dunque, di una replica del mondo reale – sia esso una persona, un ambiente, un processo produttivo, un'intera città, o un'astrazione – che ne permette non solo il monitoraggio dei fenomeni che vi avvengono, ma anche di altre importanti funzioni, quali la manutenzione preventiva, l'ottimizzazione delle prestazioni, il supporto alla progettazione e allo sviluppo. Funzioni massimizzate proprio dalla combinazione tra gemelli digitali e ricorso all'IA, che analizza i dati raccolti dal gemello digitale, identifica modelli e anomalie, fornisce previsioni utili, supporta la decisione.

Inizialmente radicati nel progresso tecnologico e scientifico del settore privato per il miglioramento della produzione e della gestione industriale (Crespi *et al.* 2023), i DT sono ormai riconosciuti come una leva strategica di sviluppo per il settore pubblico, tanto da essere inseriti tra i pilastri dell'Agenda 2030 delle Nazioni Unite per lo sviluppo sostenibile, e al centro di numerose iniziative – e finanziamenti – da parte dell'Unione europea e dei suoi Stati membri (European Commission 2023). Il tratto distintivo dei DT risiede nella loro capacità di replicare virtualmente – e dinamicamente – le caratteristiche di un'entità fisica concreta, di un sistema o di un processo attraverso un flusso di dati bidirezionale e in tempo reale. Questa caratteristica li differenzia da altre tecnologie di modellazione e simulazione dato che i DT non solo rappresentano la realtà ma ne anticipano le potenziali evoluzioni, permettendo di simulare e valutare l'impatto di eventuali cambiamenti sullo scenario corrente (Rullo 2024).

Il crescente interesse delle scienze sociali conferma il potenziale trasformativo di questa tecnologia e traccia una linea di evoluzione dell'idea stessa di azione pubblica verso nuove forme di controllo, progettazione, gestione (Musella 2024). Di fatto, "come manifestazione ultima e più progredita di quella spinta alla razionalizzazione del mondo che era stata a lungo il *telos* dell'Occidente, essi sono, e produrranno sempre più, un impulso al pensiero organizzativo. Spingono infatti a pensare alla realtà oggettiva come spazio del cambiamento, stimolando la dimensione sperimentale della conoscenza" (Musella 2024, 455).

Nella Pubblica amministrazione, i DT innovano il ruolo e i tradizionali canoni di funzionamento weberiani, prefigurando un ulteriore livello di trasformazione della Pubblica amministrazione basato sulla connessione in tempo reale tra il mondo fisico e quello virtuale (Eom 2022, 174). Sebbene vi siano preoccupazioni in merito ai rischi di controllo formale e trasparenza nella gestione dei dati, l'idea

15 Si veda per un inquadramento e alcune analisi comparate il numero monografico doppio della *Rivista di Digital Politics*, 3(2024)/1(2025) e in particolare l'introduzione di F. Musella (2024). Questi temi, inoltre, sono al centro del PRIN 2023-2025, *The Digital Twin Technologies in Contemporary Public Administration*.

principale è che attraverso i DT i servizi pubblici possano essere meglio pianificati e governati, in quanto i decisori saranno in grado di risolvere più rapidamente problemi "in vari campi, tra cui l'occupazione, le imprese, i trasporti, l'ambiente, il welfare, la sicurezza e l'architettura urbana" (Eom 2022, 181), anche in un'ottica di maggiore personalizzazione dei servizi offerti[16] (Kopponen *et al.* 2024).

In questo senso, la tecnologia DT amplifica la *policy capacity* analitica grazie all'utilizzo di vari tipi – e fonti – di dati in tempo reale e a strumenti di modellazione predittiva che non si limitano a guardare ciò che ha funzionato, ma anticipano quanto potrebbe funzionare. In questo modo, gli amministratori pubblici possono utilizzare queste informazioni per ridurre il margine di errore prima che una determinata policy venga implementata grazie alla realizzazione di scenari '*what if*' in tempo reale e sempre più precisi. Questo ventaglio di potenzialità – che rispecchia precise priorità politiche e organizzative – è particolarmente importante per ridurre il rischio di fallimenti di una determinata policy, nonché per facilitare un processo decisionale più trasparente e partecipativo, tanto da permettere di "formulare proposte altrimenti liquidate come irrealizzabili, di predisporre modifiche alle policy in vigore o di valutare l'impatto ambientale di grandi opere pubbliche, così da discutere la loro efficacia e rafforzare strategie alternative" (Micciarelli 2024; Novelli *et al.* 2025).

Ad oggi, è il contesto urbano l'arena in cui si stanno cogliendo le esperienze più mature. Su scala globale, i DT sono sempre più utilizzati per gestire in tempo reale le infrastrutture cittadine e l'ambiente urbano, con applicazioni che riguardano l'urbanistica, la mobilità urbana, la sostenibilità ambientale, la gestione delle emergenze (Corsi 2025). Ad esempio, nella pianificazione delle politiche di sostenibilità energetica, i DT possono simulare l'efficacia di diverse strategie di riduzione delle emissioni di CO2, consentendo ai governi di scegliere l'alternativa di policy più efficace. Oppure, come dimostra l'esperienza di S-Map (https://smap.seoul.go.kr/) di Seoul in Corea del Sud, i DT possono analizzare il flusso di traffico e prevedere l'effetto di nuove infrastrutture viarie o modifiche alla regolazione del traffico, consentendo di adottare misure preventive per ridurre la congestione (Eom 2022). Anche in Italia, lo studio e la sperimentazione di City Digital Twins rappresentano una realtà in continua evoluzione[17] e costituiscono il focus del PNRR Spoke 9 Digital Societies & Smart cities[18] che punta a identificare strategie e scenari che possono trasformare il modo in cui le città vengono gestite e vissute. Si tratta, dunque, di aree di innovazione pubblica che coinvolgono direttamente una molteplicità di attori nella realizzazione del progetto, tra cui Università, Centri di ricerca pubblici, autorità locali, aziende fornitrici di dati.

In un recente numero monografico della *Rivista di Digital Politics* abbiamo messo in evidenza come i DT accrescano le possibilità gestionali e di progettazione sociale "soffermandosi sempre più spesso, a diversi livelli autoritativi, sulla possibilità del cambiamento, avendone a disposizione importanti proiezioni verso il futuro" (Musella 2024, 462). Ma che soprattutto rivoluzionino la funzione di controllo pubblico, con sistemi in grado di restituire in tempo reale il comportamento di cittadini e attori collettivi ben al di là della tradizione benthamiana. Di conseguenza, abbiamo messo in luce la necessità di identificare condizioni pratiche e, ove possibile, protocolli intersettoriali che possano orientare in modo più consapevole la progettazione e la gestione dei DT nel settore pubblico. Sono evidenti, infatti, i rischi alti o inaccettabili derivanti dalla trasposizione nella dimensione pubblica di criteri ispirati a mere logiche di controllo (privato), dal momento che l'impiego dei DT può esporre a vulnerabilità sistemiche, a causa delle loro potenzialità sia di sorveglianza che di controllo predittivo dei comportamenti. Tali possibilità, se non adeguatamente governate attraverso un approccio *human-on-the-loop*,

16 Si considerino, inoltre, le implicazioni legate al ruolo delle Corporation che controllano già ampie porzioni di infrastrutture DT tra cui, ad esempio Siemens, Samsung, Azure di Microsoft, AWS. In Corea del Sud, inoltre, società come Samsung sono coinvolte nella creazione di DT per applicazioni militari. In Giappone, Fujitsu ha annunciato lo sviluppo di un Policy Twin per simulare a livello locale l'impatto di servizi sanitari (Rullo 2024).

17 In Italia le sperimentazioni di City Digital Twins hanno interessato numerose città, tra cui Bologna, Catania, Firenze, Genova, Matera, Milano, Napoli, Padova, Torino (Borriello e Fristachi 2025).

18 Per approfondire sugli obiettivi e i risultati conseguiti si veda: https://www.supercomputing-icsc.it/en/spoke-9-digital-society-smart-cities-en/.

rischiano di tradursi in violazioni sistemiche dei diritti fondamentali, imponendo una partecipazione più critica – e informata – degli attori pubblici e una riflessione critica sul design istituzionale dei DT coerente con il quadro normativo europeo. In particolare, emerge l'esigenza di adottare sistemi di individuazione e mitigazione del rischio, al fine di ridurre gli impatti sociali negativi sui diritti delle persone (Corsi 2025). È in questa zona – ancora grigia – che si deciderà se queste tecnologie emergenti diventeranno un volano di rigenerazione pubblica o l'anticamera di una delega silenziosa al dominio tecnologico privato. Solo rafforzando la governance democratica di questi processi si potrà scongiurare l'alternativa paralizzante tra resa e inerzia, tra cessione di sovranità e immobilismo istituzionale rispetto alle grandi trasformazioni in corso.

4. Un sentiero di innovazione amministrativa

Il dibattito sull'IA tende a concentrarsi sulla discussione di quanto questa sia intelligente (Crawford 2021; Messeri e Crockett 2024). Un punto importante di tale dibattito verte intorno all'idea di singolarità, guardando al modo in cui le nuove tecnologie possano superare, o sovvertire, l'intelletto umano. L'IA è invece una tecnologia culturale e sociale, al pari delle invenzioni che hanno caratterizzato la modernità, dalla burocrazia al mercato. Modifica dunque le logiche – e le capacità – degli attori coinvolti, a partire dagli individui fino alle organizzazioni complesse. Come le precedenti tecnologie sociali "permette di riorganizzarle, trasformarle e ristrutturarle in modi diversi" (Farrell *et al.* 2025; Farrell 2025). Una volta mostrate le potenzialità per la Pubblica amministrazione, dunque, bisogna pensare a come dotare il settore pubblico delle competenze necessarie per l'adozione delle nuove tecnologie digitali. Detta in altri termini raggiungere livelli di *policy capacity* sistemica che consentano il cambio di paradigma.

La crescita competenziale degli individui e dei gruppi impegnati nella Pubblica amministrazione acquista un posto di rilievo nei programmi di formazione delle nuove tecnologie a livello amministrativo (Musella e Reda 2023). Si tratta innanzitutto del piano dell'expertise (Peters 2015): quando si parla di digitale la prima constatazione attiene alla mancanza di conoscenze specialistiche e specifiche abilità che l'amministratore pubblico, formato e selezionato secondo logiche lontane dall'innovazione digitale, non può possedere. Presupporre che l'innovazione tecnologica produca cambiamenti strutturali indipendentemente dalla formazione dei burocrati equivale a trascurare ciò che rende l'amministrazione pubblica un sistema complesso (Giest e Klievink 2022). È necessario, dunque, investire sulla formazione per incrementare sia i livelli di consapevolezza in merito alle potenzialità dell'IA che quelli di apertura al cambiamento. Dinamiche, dunque, che possono portare gli amministratori a comprendere e operare nel nuovo contesto digitale ed essere i protagonisti dei processi di trasformazione digitale in corso.

D'altra parte, è proprio la diffusione dell'IA che produce un potenziamento delle *policy capacity*, nelle sue tre dimensioni principali. A partire da quella di tipo analitico: i nuovi strumenti permettono agli amministratori un ventaglio di informazioni e conoscenze indisponibili sino a un recente passato; efficaci mezzi di raccolta ed elaborazione di immense moli di dati; nuove modalità per analizzare i problemi, prevedere gli impatti delle politiche e prendere decisioni informate. Per andare alle capacità operative, con le amplissime possibilità di pianificazione e coordinamento offerte dall'algoritmizzazione del lavoro amministrativo. Con il rafforzamento, infine, della capacità di ordine politico a intervenire quando il *public management* viene orientato a ricercare nuovi modi di formulare le politiche, quando l'amministrazione entra nel pieno della definizione di nuove pratiche e relazioni.

Si aprono, dunque, sentieri di innovazione amministrativa che, tuttavia, non seguiranno traiettorie lineari o predeterminate. La *policy capacity*, infatti, non è una dotazione fissa o un pacchetto di competenze, abilità, risorse preesistenti o da distribuire una volta per tutte (Wellstead *et al.* 2024). È da intendersi, dunque, come un'infrastruttura da costruirsi attraverso la formazione proprio perché costituita da competenze, conoscenze e pratiche concrete che possono essere apprese e sviluppate in contesti formativi adeguati. In altre parole, può migliorare solo se accompagnata da investimenti nello sviluppo delle competenze, soprattutto attraverso percorsi specifici che valorizzino le scienze sociali. In particolare, è cruciale che gli amministratori si concentrino maggiormente sul loro valore aggiunto

in termini di pensiero critico rispetto all'IA e ai suoi output, nell'ambito più ampio di una trasformazione delle funzioni svolte, così da garantire livelli adeguati di comprensione in merito al potenziale e ai pericoli che ne conseguono.

Le nuove tecnologie sono state troppo spesso interpretate come la via maestra per ridurre le unità di personale della Pubblica amministrazione e trasformare ogni decisione dell'uomo in calcolo della macchina algoritmica (Karp e Zamiska 2025). Da questo punto di vista è proprio la funzione di progettazione che potrebbe essere particolarmente valorizzata in futuro. La relazione tra IA e politiche pubbliche è ancora tutta da scoprire.

Bibliografia

Agid (2025), *L'intelligenza artificiale nelle pubbliche amministrazioni. Rapporto 2025*, Roma, Agid

Agid (2024), *Strategia Italiana per l'Intelligenza artificiale*, Roma, Agid

Agid (2023), *Piano triennale per l'Informatica nella Pubblica amministrazione. Edizione 2024-2026*, Roma, Agid

Albè G., Bottini F. (2022), Consob e la "SupTech": il ruolo dell'intelligenza artificiale nell'attività di vigilanza, *Agenda Digitale,* 25 luglio

Albrecht E. (2025), *Political automation: an introduction to ai in government and its impact on citizens*, Oxford, Oxford University Press

Amoretti F., Musella F. (2012), Policy e politics del governo elettronico. L'esperienza europea, *Rivista italiana di Politiche pubbliche*, 7, n.3, pp.321-348

Armiento M.B. (2025), *Pubbliche amministrazioni e intelligenza artificiale. Strumenti, principi e garanzie*, Napoli, Editoriale Scientifica

Bobbio N. (1980), La democrazia e il potere invisibile, *Italian Political Science Review / Rivista Italiana di Scienza Politica*, 10, n.2, pp.181-203

Borghetti A. (2023), L'Intelligenza artificiale nei controlli tributari. Lotta all'evasione fiscale, *Sinappsi*, XIII, n.3, pp.98-110

Borriello G., Fristachi G. (2025), The impact of Digital twins and isomorphism on Italian municipalities, *Rivista di Digital Politics*, 5, n.1, pp.115-139

Cacciatore F. (2024), Digital citizens as public service takers: simplification and digitalization policies in Italy, in Torino F., *Digital Citizenship in The European Union Framework. Political, Economic, Sociological, and Legal Issues*, Roma, Roma Tre Press, pp.41-55

Calise M., Musella F. (2023), Digicrazia. Istruzioni per l'uso, *Rivista di Digital Politics*, 3, n.3, pp.461-480

Calise M., Musella F. (2019), *Il Principe digitale*, Roma-Bari, Laterza

Cantens T. (2025), How will the state think with ChatGPT? The challenges of generative artificial intelligence for public administrations, *AI & SOCIETY*, 40, n.1, pp.133-144

Capano G., Pritoni A., Profeti S. (2025), Does Policy Capacity Truly Matter for Governmental Effectiveness? A Conjunctural Analysis of the Quality of Governance in Italian Regions, *Governance*, 38, n.3, e70027.

Chandra Y., Feng N. (2025), Algorithms for a new season? Mapping a decade of research on the artificial intelligence-driven digital transformation of public administration, *Public Management Review* <DOI:10.1080/14719037.2025.2450680>

Corsi R. (2025), La digital governance del gemello digitale urbano in Europa: sfide etiche e compliance legale per le pubbliche amministrazioni nel governo della tecnologia, *Rivista di Digital Politics*, 5, n.1, pp.25-48

Cortés-Cediel M.E., Segura-Tinoco A., Cantador I., Bolívar M.P.R. (2023), Trends and challenges of e-government chatbots: advances in exploring open government data and citizen participation content, *Government Information Quarterly*, 40, n.4, 101877

Crawford K. (2021), *Né artificiale né intelligente. Il lato oscuro dell'IA*, Bologna, il Mulino

Crespi N., Drobot A.T., Minerva R. (a cura di) (2023), *The Digital Twin*, Cham, Springer International Publishing

Criado J.I., Sandoval-Almazán R., Gil-Garcia J.R. (2025), Artificial intelligence and public administration: Understanding actors, governance, and policy from micro, meso, and macro perspectives, *Public Policy and Administration*, 40, n.2, pp.173-184

Dalfino D. (2020), Decisione amministrativa robotica ed effetto performativo. Un beffardo algoritmo per una "buona scuola", *Questione Giustizia*, 13 gennaio

Dave P., Schiffer Z., Kelly M. (2025), Elon Musk's DOGE is working on a custom chatbot called GSAi, *Wired*, February 6

Di Giulio M., Vecchi G. (2023), Implementing digitalization in the public sector. Technologies, agency, and governance, *Public Policy and Administration*, 38, n.2, pp.133-158

Di Mascio F., Francisci G., Natalini A. (2025), *La fabbrica delle riforme amministrative*, Bologna, il Mulino

Di Mascio F., Natalini A. (2022), *Pubbliche amministrazioni. Tradizioni, paradigmi e percorsi di ricerca,* Bologna, il Mulino

Dunleavy P., Margetts H. (2023), Data science, artificial intelligence and the third wave of digital era governance, *Public Policy and Administration*, 40, n.2, pp.185-214

Engstrom D.F., Ho D.E. (2021), Artificially intelligent government: a review and agenda, in Vogl R. (ed.), *Research Handbook on Big Data Law*, Cheltenham UK, Edward Elgar, pp.57-86

Eom S.J. (2022), The emerging Digital twin bureaucracy in the 21st century, *Perspectives on Public Management and Governance*, 5, n.2, pp.174-186

European Commission (2023), *Content and Technology, Mapping EU-based LDT providers and users*, Bruxelles, European Commission- Directorate-General for Communications Networks

European Commission, Joint Research Centre (JRC) (2023), *PSTW: Public Sector Tech Watch latest dataset of selected cases*, [Dataset] PID: http://data.europa.eu/89h/e8e7bddd-8510-4936-9fa6-7e1b399cbd92

Farrell H. (2025), AI as Governance, *Annual Review of Political Science*, 28, n.1, pp.375-392

Farrell H., Gopnik A., Shalizi C., Evans J. (2025), Large AI models are cultural and social technologies, *Science*, 387, n.6739, pp.1153-1156

Ferrarese M.G. (2022), *Poteri nuovi*, Bologna, il Mulino

Filgueiras F. (2025), *Artificial intelligence and public policy: disrupting policy sciences*, Cambridge, Cambridge University Press

Giest S.N., Klievink B. (2022), More than a digital system: how AI is changing the role of bureaucrats in different organizational contexts, *Public Management Review,* 26, n.2, pp.379-398

Haesevoets T., Verschuere B., Van Severen R., Roets A. (2024), How do citizens perceive the use of Artificial Intelligence in public sector decisions?, *Government Information Quarterly*, 41, n.1, 101906

Howlett M. (2015), Policy analytical capacity. The supply and demand for policy analysis in government, ***Policy and Society***, 34, nn.3-4, pp.173-182

Howlett M., Giest S., Mukherjee I., Taeihagh A. (2025), New policy tools and traditional policy models: better understanding behavioural, digital and collaborative instruments, *Policy Design and Practice*, 8, n.1, pp.121-137

Karp A.C., Zamiska N.W. (2025), *The Technological Republic: Hard Power, Soft Belief, and the Future of the West*, New York, Crown Currency

Khan R., Hussain F. (2024), Assessing policy capacity and policy effectiveness: A comparative study using sustainable governance indicators, *European Policy Analysis*, 10, n.4, pp.575-603

Khanal S., Zhang H., Taeihagh A. (2025), Why and how is the power of Big Tech increasing in the policy process? The case of generative AI, *Policy and Society*, 44, n.1, pp.52-69

Kim S., Wellstead A. M., Heikkila T. (2023), Policy capacity and rise of data-based policy innovation labs, *Review of Policy Research*, 40, n.3, pp.341-362

Kopponen A., Hahto A., Villman T., Kettunen P., Mikkonen T., Rossi M. (2024), Personalised public services powered by AI. The citizen digital twin approach, in *Research Handbook on Public Management and Artificial Intelligence*, Cheltenham, Edward Elgar Publishing, pp.170-186

Larsen A.G., Følstad A. (2024), The impact of chatbots on public service provision: A qualitative interview study with citizens and public service providers, *Government Information Quarterly*, 41, n.2, 101927

Licata G.B. (2024), Intelligenza artificiale e contratti pubblici: problemi e prospettive, *Rivista interdisciplinare sul diritto delle amministrazioni pubbliche*, n.2, pp.30-63

Margetts H., Dorobantu C., Bright J. (2024), How to build progressive public services with data science and artificial intelligence, *The Political Quarterly*, 95, n.4, pp.653-662

Meijer A., Lorenz L., Wessels M. (2021), Algorithmization of bureaucratic organizations: using a practice lens to study how context shapes predictive policing systems, *Public Administration Review*, 81, n.5, pp.837-846

Mergel I., Dickinson H., Stenvall J., Gasco M. (2024), Implementing AI in the public sector, *Public Management Review* <https://doi.org/10.1080/14719037.2023.2231950>

Messeri L., Crockett, M.J. (2024), Artificial intelligence and illusions of understanding in scientific research, *Nature*, 627, n.8002, pp.49-58

Micciarelli G. (2024), Architettura e logica politica dei Digital twins: un approccio critico e prefigurativo, *Rivista di Digital Politics*, 4, n.3, pp.513-535

Misuraca G., van Noordt C. (2020), *Overview of the use and impact of ai in public services in the Eu,* Luxembourg, Publications Office of the European Union <DOI:10.2760/039619>

Misuraca, G., Rossel, P., Sibal, P. (2024), Mastering AI governance in the public sector, in *The Routledge International Handbook of Public Administration and Digital Governance*, London, Routledge, pp.338-353

Mukherjee I., Coban M.K., Bali A.S. (2021), Policy capacities and effective policy design. A review, *Policy Sciences*, 54, n.2, pp.243-268

Musella F. (2024), Il futuro dei Digital twins, *Rivista di Digital Politics*, 4, n.3, pp.453-464

Musella F. (2022), Digital regulation: come si cambia la Pubblica amministrazione, *Rivista di Digital Politics*, 2, nn.1-2, pp.3-32

Musella F. (2021), Amministrazione 5.0, *Rivista di Digital Politics*, 1, n.1, pp.95-112

Musella F., Reda V. (2023), The 'great re-skilling' e la formazione digitale per la PA, *Sinappsi*, XIII, n.3, pp.6-19

Natalini A. (2022), Come il passato influenza la digitalizzazione delle amministrazioni pubbliche, *Rivista trimestrale di diritto pubblico*, n.1, pp.95-116

Novelli C., Sanchez-Vaquerizo J.A., Helbing D., Rotolo A., Floridi L. (2025), A replica for our democracies? on using digital twins to enhance deliberative democracy, *AI & Society* <https://doi.org/10.1007/s00146-025-02511-7>

Osborne D., Gaebler T. (1992), *Reinventing government: how the entrepreneurial spirit is transforming the public sector*, New York, Perseus Books

Pencheva I., Esteve M., Mikhaylov S.J. (2020), Big Data and AI–A transformational shift for government: So, what next for research?, *Public Policy and Administration*, 35, n.1, pp.24-44

Peters B.G. (2015), Policy capacity in public administration, *Policy and Society*, 34, n.3-4, pp.219-228

Radu R. (2021), Steering the governance of artificial intelligence: national strategies in perspective, *Policy and Society*, 40, n.2, pp.178-193

Rangone N., Megale L. (2025), Risks Without Rights? The EU AI Act's Approach to AI in Law and Rule-Making, *European Journal of Risk Regulation* <DOI:10.1017/err.2025.13>

Rizzo C., Gesualdo F., Lanfranchi B., Armeni R., Rota M.C., Filia A. (2021), Una piattaforma europea per il monitoraggio delle conversazioni sui vaccini su web e social network, *Bollettino epidemiologico nazionale*, 2, n.2, pp.7-15

Rullo L. (2024), Paths of Digital twins in the public sector: a systematic review of the social sciences literature, *Rivista di Digital Politics*, 4, n.3, pp.631-653

Salah M., Abdelfattah F., Al Halbusi, H. (2023), Generative artificial intelligence (ChatGPT & Bard) in public administration research: a double-edged sword for street-level bureaucracy studies, *International Journal of Public Administration* <DOI:10.1080/01900692.2023.2274801>

Santoriello S.C. (2025), Whose Reality? Consent Boundaries and Free Speech Arguments in the Politics of Generative AI, *Politikon: The IAPSS Journal of Political Science*, 59, n.2, pp.29-56

Sgueo G. (2024), *La democrazia Migliore. Tecnologie che trasformano il potere*, Soveria-Mannelli, Rubbettino

Smuha N.A. (2025), The use of algorithmic systems by public administrations practices, challenges and governance frameworks, in Smuha N.A., *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence*, Cambridge, Cambridge University Press, pp.383-410

Tangi L., Combetto M., Hupont Torres I., Farrell E., Schade S. (2024), *The potential of generative AI for the public sector: current use, key questions and policy considerations*, JRC Research Reports, JRC139825, European Commission, Ispra

The Economist (2024), How businesses are actually using generative AI, *The Economist*, February 29th

Todaro D. (2024), *The Use of Artificial Intelligence in the Public Sector in Shanghai: Ambition, Capacity and Reality*, Singapore, Springer Nature Singapore

Torchia L. (2021), *L'amministrazione presa sul serio e l'attuazione del Pnrr*, AstridPolicy Brief-Proposte per la ripresa, Roma, Astrid

Urban J. (2025), *Policy making in the era of artificial intelligence*, Insight paper, London, Institute for government

Valle-Cruz D., Criado J.I., Sandoval-Almazán R., Ruvalcaba-Gomez E.A. (2020), Assessing the public policy-cycle framework in the age of artificial intelligence. From agenda-setting to policy evaluation, *Government Information Quarterly*, 37, n.4, pp.1-12

van Dijck J., van Es K., Helmond A., van der Vlist F. (2025), *Governing the Digital Society: Platforms, Artificial Intelligence, and Public Values*, Amsterdam, Amsterdam University Press

van Noordt C., Medaglia R., Tangi L. (2025), Policy initiatives for Artificial Intelligence-enabled government: An analysis of national strategies in Europe, *Public Policy and Administration*, 40, n.2, pp.215-253

van Ooijen C., Ubaldi B., Welby B. (2019), *A data-driven public sector: Enabling the strategic use of data for productive, inclusive and trustworthy governance*, OECD Working Papers on Public Governance, Paris, OECD Publishing

VanDerHorn E., Mahadevan S. (2021), Digital Twin: generalization, characterization and implementation, *Decision support systems* <DOI 1016/j.dss.2021.113524>

Vogl T.M., Seidelin C., Ganesh B., Bright J. (2020), Smart technology and the emergence of algorithmic bureaucracy: artificial intelligence in UK local authorities, *Public Administration Review*, 80, n.6, pp.946-961

Wellstead A.M., Mechling S.M., Carter A., Gofen A. (2024), Artificial intelligence possibilities to improve analytical policy capacity: the case of environmental policy innovation labs and sustainable development goals, *Policy Design and Practice*, 7, n.4, pp.456-467

Wirtz B.W., Müller W.M. (2019), An integrated artificial intelligence framework for public management, *Public Management Review*, 21, n.7, pp.1076-1100

Wirtz B.W., Weyerer J.C., Kehl I. (2022), Governance of artificial intelligence: A risk and guideline-based integrative framework, *Government Information Quarterly*, 39, n.4, pp.1-17

Woo J. (2021), *Capacity-Building and Pandemics. Singapore's Response to Covid-19*, Singapore, Palgrave Macmillan

Wu X., Howlett M., Ramesh M. (eds.) (2018), *Policy Capacity and Governance. Assessing Governmental Competences and Capabilities in Theory and Practice*, Cham, Palgrave Macmillan

Wu X., Ramesh M., Howlett M. (2015), Policy capacity. A conceptual framework for understanding policy competences and capabilities, *Policy and Society*, 34, n.3-4, pp.165-171

Yar M.A., Hamdan M., Anshari M., Fitriyani N.L., Syafrudin M. (2024), Governing with intelligence: the impact of artificial intelligence on policy development, *Information*, 15, n.9, pp.1-17

Fortunato Musella
fortunato.musella@unina.it
Professore ordinario di Scienza politica presso l'Università degli Studi di Napoli Federico II, dove è anche Direttore di Federica WebLearning, Centro per l'Innovazione e la diffusione della formazione a distanza, e delegato del Rettore dell'Università Federico II per le attività didattiche. È Responsabile scientifico Digital Education Hub ALMA - Advanced Learning Multimedia Alliance e fondatore e membro del Consiglio di amministrazione della Scuola di management pubblico della Federico II (SPM). È responsabile scientifico del progetto di ricerca The Digital Twin Technologies in Contemporary Public Administration. Tra le sue recenti pubblicazioni: con M. Calise, *Il Principe digitale*, Laterza, 2019; *Il governo in Italia*, il Mulino, 2019; *Monocratic Government*, De Gruyter, 2022.

Luigi Rullo
luigi.rullo@unina.it
Ricercatore di Scienza politica presso l'Università degli Studi di Napoli Federico II e docente di Innovazione politica digitale presso il Dipartimento di Scienze sociali della stessa università. È membro dell'editorial management della *Rivista di Digital Politics* (il Mulino) e Book review Editor della rivista *Contemporary Italian Politics* (Routledge). Tra le pubblicazioni più recenti si segnalano (con F. Musella), The personalization of government: concept and comparative analysis, *Democratization,* 2024; Paths of Digital twins in the public sector: A systematic review of the social sciences literature, *Rivista di Digital Politics,* 2024.

Intelligente e artificiale: una tecnologia organizzativa da regolare

Giorgio Gosetti
Università degli Studi di Verona
Marco Peruzzi
Università degli Studi di Verona

I processi di innovazione nel lavoro sono sempre più legati all'introduzione dell'Intelligenza artificiale. Così come la digitalizzazione, anche l'Intelligenza artificiale sta generando un forte impatto sul mondo del lavoro, in termini quantitativi, ma soprattutto qualitativi. Infatti, sono ridisegnarti processi e contenuti del lavoro e di conseguenza anche le condizioni della qualità della vita lavorativa. In questo scenario diventa quanto mai necessario riflettere sulla progettazione organizzativa supportata dalle tecnologie e soprattutto sui sistemi di regolazione a diversi livelli.

Innovation processes in the workplace are increasingly driven by the introduction of Artificial intelligence. Like digitisation, artificial intelligence is having a significant impact on the world of work, not only in quantitative terms but, above all, in qualitative ones. Work processes and content are being reshaped, and so are the conditions affecting the quality of working life. In this context, it is more necessary than ever to reflect on technology-supported organisational design and, above all, on regulatory frameworks at various levels.

DOI: 10.53223/Sinappsi_2025-02-10

Citazione
Gosetti G., Peruzzi M. (2025), Intelligente e artificiale: una tecnologia organizzativa da regolare, *Sinappsi*, XV, n.2, pp.128-155

Parole chiave
Intelligenza artificiale
Organizzazione del lavoro
Tutele del lavoro

Keywords
Artificial intelligence
Work organisation
Labour protection

Introduzione

I processi di innovazione nel lavoro sono sempre più legati all'introduzione dell'Intelligenza artificiale (IA). Così come la digitalizzazione, anche l'IA sta generando un forte impatto sul mondo del lavoro, in termini quantitativi, ma soprattutto qualitativi. Infatti, vengono ridisegnati processi e contenuti del lavoro e, di conseguenza, anche le condizioni della qualità del lavoro e della vita lavorativa. In questo scenario è quanto mai necessario riflettere sulla progettazione organizzativa supportata dalle tecnologie e soprattutto sui sistemi di regolazione a diversi livelli.

A partire da queste premesse il saggio si articola in due parti. Nella prima, un approccio sociologico ai temi del lavoro cerca di porre in evidenza i tratti salienti dell'IA e le implicazioni sul lavoro, con particolare attenzione agli aspetti organizzativi e di qualità della vita lavorativa. Nella seconda, improntata a un approccio giuslavoristico, è sviluppata una riflessione sull'architettura regolativa chiamata a governare il fenomeno e sul tipo e grado

Il testo è frutto di un lavoro di riflessione congiunta fra i due Autori. Ai fini formali e per le rispettive competenze disciplinari vanno comunque attribuiti a Giorgio Gosetti il paragrafo 1 e a Marco Peruzzi il paragrafo 2.

di tutele che può fornire al lavoratore, sul piano individuale e collettivo. Le due sezioni di questo contributo sono attraversate da un aspetto che le accomuna, vale a dire la prospettiva che intende leggere l'IA specificatamente come una *variabile organizzativa*. La quinta rivoluzione tecnologico/industriale va considerata specificatamente per le implicazioni che può avere sulla vita lavorativa, cogliendo le connotazioni che la differenziano dalle precedenti. Siamo infatti di fronte a una tecnologia di quinta generazione che sta producendo effetti di profondo mutamento nelle forme organizzative, nei meccanismi operativi e nei contenuti del lavoro, sconfinando in un mutamento antropologico. Agisce sul modo nel quale intendiamo il lavoro. È del tutto evidente che uno scenario di mutamento di tale portata ci richiede di considerare forme e contenuti della regolazione.

1. Digitalizzazione e Intelligenza artificiale

Iniziamo questa prima parte da una frase che apre e chiude il cerchio delle nostre riflessioni: "il rischio al quale ci espone l'IA, come ogni tecnologia ad ampio spettro, è quello di trasformare gradualmente il mondo al punto da diventare sempre più indispensabile per il suo funzionamento" (Adler 2024, 375). Proviamo a riempire di contenuto il cerchio perimetrato da questa frase, avendo sullo sfondo quel "principio di moderazione" proposto da Adler, una sorta di "forma di sussidiarietà" che si traduce nella prospettiva di "fare appello all'Intelligenza artificiale solo quando il suo contributo netto è positivo" (*Ibidem*).

Elementi per definire l'IA

In un lavoro comparso recentemente sempre su questa rivista (Canal *et al.* 2024), proponendo un'analisi a partire dai dati della V Indagine Inapp sulla Qualità del lavoro in Italia (QdL), eravamo giunti alla conclusione che il "rapporto uomo-macchina, alla base di tanta letteratura che ha studiato le condizioni lavorative, va riproblematizzato alla luce dei recenti sviluppi delle innovazioni tecnologiche". Le analisi dei dati ci hanno posto davanti a differenti profili di lavoratori legati al diverso tipo di qualificazione e specializzazione. Infatti "l'utilizzo della tecnologia pervasiva può (…) diventare un'occasione per generare *capabilities*, per declinare concretamente un saper apprendere e un saper agire dentro i luoghi di lavoro", proprio perché interviene sul tipo e sul livello di qualificazione e specializzazione del lavoratore. Il contenuto del lavoro legato alla tecnologia digitale è infatti direttamente riferibile alle possibilità per il lavoratore di un'azione autonoma e di un controllo sul proprio lavoro. Fuori da ogni determinismo tecnologico, il grado di qualificazione del lavoratore appare comunque segnato dal tipo di tecnologia utilizzata: "specializzazioni in tecnologie hardware si legano a profili simili a quelli dei lavoratori non digitali e, addirittura, in alcuni casi paiono più fragili (ad esempio per la maggiore presenza di contratti non standard); al contrario, la specializzazione in tecnologie software si associa a migliori profili occupazionali. (…) I lavoratori specializzati nell'utilizzo di strumenti produttivi di tipo software beneficiano di un ritorno diretto e complessivo in termini di qualità del lavoro, rispetto a coloro non interessati da strumenti digitali. Al contrario, l'utilizzo di tecnologie hardware non si associa a effetti positivi, anzi genera ricadute negative in termini ergonomici". Una differenza che ritroviamo anche esaminando la percezione di essere esposti a un rischio di "obsolescenza tecnologica" o di "dissonanza" fra attività svolta e aspirazioni lavorative (Canal *et al.* 2024, 81-82). Tutti elementi che rinnovano l'invito a studiare e monitorare gli effetti quantitativi (sulle dinamiche occupazionali) e qualitativi (sulla qualità del lavoro e della vita lavorativa) generati dall'introduzione delle innovazioni tecnologiche.

Il percorso che intendiamo operare in questo saggio parte da queste premesse, ma fa proprio un ulteriore presupposto: rispetto alla recente ondata di innovazioni nel mondo del lavoro legate alla digitalizzazione, siamo di fronte a una nuova fase di mutamento, che ruota e ruoterà sempre più attorno all'IA. Le tecnologie che possiamo collocare sotto l'etichetta di IA si muovono in continuità con quelle della digitalizzazione ormai matura, ma ci pare evidenzino alcuni tratti distintivi rispetto ad essa. Tratti che ci portano a parlare ormai di una quinta rivoluzione tecnologica: dopo il vapore che muove i sistemi di produzione e trasporto, l'elettricità che alimenta le catene di montaggio, l'automazione e l'informatica che generano sistemi di produzione centrati sul trattamento di informazioni, la digitalizzazione che connette e pervade i sistemi di produzione e di vita, siamo arrivati a un *nuovo mondo* segnato dall'IA. Un mutamento da affrontare in termini interdisciplinari sotto il profilo analitico

e regolatorio (Gosetti *et al.* 2024), che riguarda un passaggio antropologico profondo. Nelle prossime pagine, inizialmente partiremo da alcuni aspetti definitori e di inquadramento sociologico del tema, per poi trovare elementi di discussione sotto il profilo giuridico e della regolazione a diversi livelli. Il terreno unificante rimane quello del lavoro, nelle sue diverse dimensioni.

Nel cercare elementi per una definizione di Intelligenza artificiale, ci viene incontro quanto recentemente stabilito dall'Unione europea (European Commission 2025): "*'AI system'means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments*". Partendo dalla varietà di elementi che caratterizza l'IA e che è alla base della difficoltà a produrre una definizione esaustiva di IA, le linee tracciate dall'Unione europea sostengono che "*the notion of an 'AI system'should be clearly defined while providing the flexibility to accommodate the rapid technological developments in this field. The definition of an AI system should not be applied mechanically; each system must be assessed based on its specific characteristics*". Nello specifico, a corollario della definizione, sono individuati sette elementi basilari: *"(1) a machine-based system; (2) that is designed to operate with varying levels of autonomy; (3) that may exhibit adaptiveness after deployment; (4) and that, for explicit or implicit objectives; (5) infers, from the input it receives, how to generate outputs* (6) *such as predictions, content, recommendations, or decisions (7) that can influence physical or virtual environments*". Ci riferiamo quindi a un *sistema*, che richiama immediatamente una logica di parti in forte interazione, in grado di agire a differenti livelli di *autonomia* operativa, di *adattarsi*, modulando l'impiego di risorse a seconda degli *obiettivi* da perseguire, essendo in grado di *rilevare* e *trattare* informazioni, di formulare *previsioni* a partire da esse e, soprattutto, di prendere *decisioni* immediatamente *influenti* sull'ambiente fisico e virtuale. Un sistema che possiamo analizzare guardando al suo ciclo di vita, quindi distinguendo due fasi, quella di '*pre-deployment or building*' da quella di '*post-deployment or use*'. I sette elementi ricordati poco sopra compaiono a diversa gradazione nelle due fasi e, come appare evidente, una definizione di IA così formulata risponde anche all'esigenza dell'Unione europea di includere un'ampia gamma di sistemi e di creare le premesse per azioni di governo.

Ma agli elementi di questa prima disamina, che potremmo qualificare di ordine politico-istituzionale, possiamo affiancarne altri, ricorrendo alla letteratura di carattere sociologico che sempre più spesso in questa fase sta discutendo il tema dell'IA. Sceglieremo una lettura che ci porta nel paragrafo successivo a tracciare alcune traiettorie interpretative provvisorie. Ci pare di poter dire che, per ora, prevalgono più le domande delle risposte.

Possiamo innanzitutto partire da una considerazione generale sostenendo che "l'AI consente alle macchine di apprendere, pensare e agire in base a processi squisitamente umani, di ordine razionale, sollevando problemi tecnici, filosofici ed etici. (…) Mettere in grado una macchina di fissare i propri obiettivi e perseguirli, monitorare i risultati delle proprie azioni e modificarli in base al comportamento imprevedibile di altre entità attive nello stesso ambiente, apprendere automaticamente attraverso l'esperienza, in modo da migliorare il proprio rendimento, comprendere il linguaggio degli esseri umani, rispondere a domande, eseguire ordini, analizzare e tradurre testi rappresentano traguardi strepitosi se si pensa che sono stati raggiunti in poche decine di anni" (De Masi 2018, 658-659). L'idea di una macchina che opera in autonomia è spesso prevalente: "Non è facile precisare in cosa consista tale autonomia, ma ora requisito minimo sarebbe che l'ideatore umano non sia indotto a precisare ogni fase dell'esecuzione del programma. (…) L'IA (…) non ha un rapporto diretto con l'intelligenza umana; (…) assembla, pezzo per pezzo, una cassetta degli attrezzi che s'iscrive nella continuità dei processi di automatizzazione e di amplificazione delle capacità umane" (Adler 2024, 26-34). Si tratta quindi di "pratiche tecniche e sociali, istituzioni e infrastrutture, politica e cultura. La ragione computazionale e il lavoro incarnato sono profondamente interconnessi: i sistemi di IA riflettono e producono relazioni sociali e comprensioni del mondo" (Crawford 2021, 15). Seppur autonoma, l'IA è radicata, esprime un disegno di società e di governo politico dei processi (sociali, economici ecc.). Come si vedrà più oltre potremmo dire che è anche incorporata.

L'IA è strettamente legata ai dati che riesce a reperire e trattare: "Questa è una differenza sostanziale col modello industriale, dove c'era una tensione tra quantità e qualità. L'intelligenza delle macchine sintetizza tale tensione, visto che raggiunge il suo massimo potenziale qualitativo solo approssimandosi alla totalità" (Zuboff 2019, 106). Cerca di trattare dati generando "*prodotti predittivi* pensati per intuire cosa sentiremo, penseremo e faremo" (Zuboff 2019, 106). Il salto di qualità negli ultimi anni è stato quello del passaggio all'IA generativa, a sistemi in grado di produrre immagini, testi e altro: "La prima ondata di GAI (*Generative Artificial intelligence*) si concentra soprattutto sulla possibilità di conversare servendosi di una lingua naturale. Chiamati '*Large Language Models*' (LLM), questi programmi hanno già dimostrato la loro sbalorditiva capacità di eseguire una serie di compiti con performance sovrumane. (…) L'AI generativa potrebbe innescare un nuovo scarto culturale, spostando il focus sulle macchine, permettendoci di utilizzare la potenza dell'intelligenza sintetica come strumento per accelerare il progresso (…). Le AI generative scardinano i nostri luoghi comuni sui computer. Dall'alba dell'era elettronica, i computer sono stati considerati una sorta di controparte misteriosa degli esseri umani: dotati di precisione infallibile e velocità smodata, freddi e calcolatori, privi di giudizi morali e di *savoir faire* sociale. Oggi, le GAI danno spesso sfoggio di impeccabili capacità interpersonali, di empatia e compassione, per quanto per ora siano prive di esperienze effettive nel mondo reale" (Kaplan 2024, 11-13).

L'IA è quindi un sistema in grado di (a) *rilevare* dati, leggendo il contesto nel quale opera, (b) *conservare* dati, immagazzinando anche grandi quantità di informazioni senza perdere memoria di quanto avviene nel contesto, (c) *elaborare* dati, consentendo di interpretare fenomeni che avvengono nel contesto, (d) *proiettare*, connettendo passato, presente e futuro a partire dai dati a disposizione, per prefigurare scenari possibili, (e) *decidere*, passando all'azione e generando conseguenze delle azioni intraprese, (f) *monitorare*, tenendo sotto controllo gli sviluppi del processo decisionale, (g) *valutare*, producendo analisi sugli effetti del processo decisionale e generando informazioni che vengono immagazzinate e consentono di innescare nuovi processi decisionali, (h) *apprendere*, ridefinendo i propri processi operativi a partire da elementi raccolti nel corso dell'azione e dagli esisti dei processi.

Ci pare di poter dire che il vero e proprio salto di qualità è dato dal fatto che, a un certo punto del suo percorso, l'IA è stata in grado contemporaneamente di (a) fornire risposte, (b) 'automatizzare task', attività altrimenti eseguite da esseri umani, (c) generare prodotti di vario tipo. L'IA generativa, in particolare, "è costituita da modelli che non operano sulla base di programmi ma che vengono addestrati con miliardi di parametri" che in questo modo imparano e assolvono una quantità elevata di funzioni. Questo non significa che sia in grado di *comprendere* completamente il contesto, di fare esperienze sensoriali dirette, sviluppare una creatività intuitiva: "l'area della creatività, dell'invenzione sorprendente, dell'apertura di nuove strade è ancora appannaggio degli esseri umani". Secondo molti autori, quindi, l'IA vede (per ora?) un limite significativo rispetto a quella umana nella "comprensione profonda dei processi umani, nella creatività, nell'empatia" (Butera e De Michelis 2024, 27-31). Si tratta di una tecnologia che è fondata sulla quantità e qualità dei dati a disposizione, ha continuamente bisogno di dati per agire e addestrarsi, secondo una logica di *machine learning*.

Uno dei capitoli più interessanti del dibattito attuale sull'evoluzione dell'IA è quello che ruota attorno al *deep learning*, l'apprendimento profondo, legato alla simulazione di reti di neuroni che per gradi sviluppano processi di apprendimento, riconoscendo immagini e linguaggi. Coloro che sviluppano reti neurali adottano "il cervello come modello di massima per progettare sistemi intelligenti". Più strati di neuroni virtuali dentro la rete definiscono la sua "profondità" e quindi il livello di capacità di "apprendere qualcosa del mondo che ha intorno" (Bengio 2016, 24-26; Mitchell 2022). Più in specifico potremmo dire che "un 'grande modello del linguaggio' (*Large Language Model*, LLM) è, essenzialmente, una *rete neurale* di grandi dimensioni", denominata in questo modo perché costituisce "un sistema artificiale che simula la struttura del cervello", composto da "unità di elaborazione ('nodi') connesse tra loro, come i neuroni; le connessioni tra i nodi simulano le sinapsi che, nel cervello, consentono la comunicazione tra i neuroni. I nodi della rete sono organizzati in 'strati'; ciascun nodo appartiene a uno e un solo strato. Ogni rete è fatta di uno strato di input, uno di output, e

uno o più strati intermedi (detti 'nascosti')" (Marconi 2025, 101).

Il *deep learning* è un "metodo tra i molti nel campo del *machine learning*", un sottocampo dell'IA che intende le macchine capaci di imparare e apprendere dalle "proprie esperienze" (Mitchell 2022, 9). L'infrastruttura alla base è fatta di 'soluzioni hardware e software' espressamente orientate a realizzare e ottimizzare le reti neurali profonde (Ford 2022). Quello degli algoritmi è comunque un comportamento che "non ha nulla a che vedere con quello del cervello umano". Le reti neurali profonde "per il momento, (…) non hanno alcuna rappresentazione del mondo né della sua organizzazione spaziale né delle relazioni causali. (…) Un modello generativo di linguaggio produce frasi (sequenze di parole più probabili) che non hanno senso per lui. Non c'è intenzionalità nel suo testo, anche se noi lettori tendiamo ad attribuirgliene, abituati come siamo a testi che fino a oggi erano sempre stati prodotti da esseri umani dotati di intenzione" (Mézard 2024, 43).

Gli algoritmi del *deep learning*, inteso appunto come apprendimento profondo, sono in grado di 'lavorare su dati non strutturati', 'imparano da sé'**.** Mentre l'apprendimento automatico del *machine learning* opera a partire da dati strutturati, organizzati ed elaborati, nel caso del *deep learning* "testi e immagini sono analizzati direttamente", non serve un intervento preordinatore dell'essere umano, e le loro peculiarità sono messe a confronto con altri dati "in un progressivo affinamento. Si chiamano processi di ottimizzazione, basati su reti neurali molto complesse e stratificate" che implicano un certo costo in termini di memoria e potenza di calcolo (Perilli 2025, 64). Andando in questa direzione si arriva a quella che viene definita 'intelligenza computazionale', vale a dire la capacità di un computer di imparare a eseguire compiti associando dati disponibili e 'osservazione sperimentale', pensando a un apprendimento esperienziale che vede la macchina capace di "modificare le sue proprie istruzioni per adattarsi alle diverse situazioni" (Perilli 2025, 98). Le macchine imparano e cambiano rispondendo agli stimoli che arrivano dall'esterno, si evolvono gradualmente. In sostanza, il *deep learning* parte dal presupposto che non sia "necessario" e "opportuno dire alla macchina cosa deve fare in dettaglio" e che i sistemi "possano apprendere ad apprendere, sviluppando una propria capacità a partire da indicazioni generiche – algoritmi – riguardanti il modo in cui esaminare i dati". Le reti neurali addestrate a partire da esempi sono "capaci di apprendere probabilisticamente, senza supervisione, a partire da dati la cui qualità è ritenuta irrilevante" (Varanini 2024, 163).

Non va comunque dimenticato, per inquadrare ulteriormente questo passaggio, che "per capire come l'IA sia fondamentalmente politica, dobbiamo andare oltre le reti neurali e il riconoscimento di modelli (*pattern recognition*) per chiederci invece, che cosa viene ottimizzato, per chi e chi è che decide" (Crawford 2021, 17). Dobbiamo quindi declinare la chiave di lettura sociologica applicata alle dinamiche dell'economia e del lavoro includendo anche variabili politiche, in grado di aiutarci a capire le micro, meso e macro-strategie che si stanno giocando attorno all'IA. Una declinazione di strategie che avviene contemporaneamente al livello dei luoghi di lavoro (micro) che stiamo continuamente ridisegnando (fino a farci pensare se per alcune attività esista ancora un 'luogo' del lavoro), dei territori (meso), sempre più a perimetrazione flessibile a seconda dello sviluppo delle catene del valore e, necessariamente, delle relazioni globali (macro). L'IA entra direttamente in quella continua stratificazione spazio-temporale che riguarda i cosiddetti processi di glocalizzazione che connotano il mutamento frequente dei tre livelli e della loro fusione continua.

Linee interpretative dell'IA

Ripercorrendo quanto abbiamo sostenuto finora, pensiamo sia possibile individuare alcune linee interpretative dell'IA in quanto tecnologia che agisce direttamente su organizzazione e contenuto del lavoro, e più specificatamente sulla qualità delle condizioni lavorative. Proveremo nelle prossime pagine a proporre alcune traiettorie partendo da *parole chiave* che ci aiutano a sintetizzare.

(a) Lavoratori

Un modo per interpretare il ruolo dell'IA è quello di andare a comprendere quanto sia in grado di modificare il lavoro umano, magari potenziandolo, generando quindi 'lavoratori arricchiti' attraverso la tecnologia (associando, ad esempio, i tradizionali *job enlargement* e *job enrichment*), oppure tenda a sostituire del tutto o in parte il lavoro delle persone. Più che ampliare le competenze e le capacità di

azione del lavoratore o magari sostituirlo, mettendo a disposizione informazioni e riconfigurando i posti di lavoro sotto il profilo organizzativo e contenutistico, l'IA è in grado di ridefinire radicalmente il lavoro, reinventarlo, ridisegnare alla base i processi di creazione del valore. Andando in questa direzione finirebbe appunto per ridefinire quello che finora abbiamo chiamato *lavoro*, inteso come attività associata a una persona impegnata nello svolgere operazioni per produrre beni o servizi dentro meccanismi operativi organizzati e in siti identificabili, sostituendolo con attività non formalizzate che generano valore (economico), ma difficilmente assimilabili appunto alla nostra tradizionale definizione di lavoro.

Dentro questa traiettoria trasformativa, per ora, si collocano quei mutamenti di organizzazione e soprattutto di contenuto del lavoro che stanno innescando un superamento della tradizionale distinzione fra colletti bianchi e colletti blu. Si stanno imponendo i cosiddetti *new collar*, differentemente colorati, che, anche quando svolgono compiti una volta associabili, ad esempio, ai colletti blu, si trovano ora a trattare informazioni come i colletti bianchi, o a essere affiancati dai – e integrati nei – sistemi di IA che annullano *alla fonte* la precisa distinzione bianco-blu, omogeneizzando colletti e divise. La diffusione dell'IA nel lavoro costituisce sicuramente un ulteriore fattore di spinta verso quella ibridazione alla base della società dei lavori (Gosetti 2024), quindi verso una combinazione di competenze tecniche e *soft skills*, contenuti scientifici riferibili alle aree tecnologiche tradizionali e alle aree umanistiche (integrando discipline), competenze di specializzazione orizzontale e verticale. Competenze che mutano nel tempo, seguendo il corso di vita dell'organizzazione e della persona. Gli aspetti tecnologici di cui ci stiamo occupando, infatti, devono anche interessare una riflessione sui passaggi generazionali e sull'adattamento delle competenze ai cicli di vita lavorativa. I nuovi contesti lavorativi richiederanno forse sempre più una capacità sistemica di lettura del proprio contesto lavorativo, una leadership legata all'IA che sa usare anche la IA per generare collaborazione e condivisione nei luoghi di lavoro. L'IA sarà sempre più orientata a supportare la selezione del personale, la formulazione della strategia organizzativa e la valutazione delle performance. Basti pensare, solo per fare un esempio, ai dispositivi che vengono impiegati nel mondo del calcio per i processi di scouting (individuando i 'talenti'), elaborare strategie e tattiche (prima della partita e durante il suo svolgimento), studiare i vari aspetti del giocatore nel corso della prestazione sportiva e proporre decisioni consequenziali, individuare percorsi personalizzati di sviluppo delle competenze.

Nei prossimi anni il lavoro sarà diverso da quello che abbiamo visto finora e si apriranno spazi per le diversità (potremmo forse dire la neurodiversità). Dovremo studiare pertanto quanto l'IA sarà in grado di creare 'ambienti' inclusivi, di rendere 'abitabile' il lavoro nelle sue diverse dimensioni. L'IA non sarà solamente sostitutiva o integrativa/complementare/supportiva, ma trasformerà il lavoro, diventerà un generatore di lavoro diverso, che ci chiederà nuovi lavoratori: *model & prompt engineers*, *data curators & trainers*, addestratori delle macchine, specialisti in etica applicata alle tecnologie e così via. Essendo una tecnologia non in grado di mettere in campo l'intelligenza emotiva, è del tutto probabile che vi sia bisogno di lavoratori, o forse ancor più di lavoratrici (dato che spesso queste competenze vengono attribuite alla componente femminile), che dispongano di questo tipo di competenze, così come di creatività e capacità analitiche. Ma, per le informazioni che abbiamo, siamo sicuramente in presenza di una tecnologia che richiede ancora molto lavoro umano, non di rado precario, sottopagato, ripetitivo, che facilita i processi dell'IA.

(b) Autonomia

Spesso quando si parla di IA si cita l'autonomia come elemento distintivo (e poco sopra vi abbiamo fatto ricorso cercando una definizione di IA). L'autonomia è, ad esempio, quella che nel lavoro vede l'incontro fra robotica e IA, quando siamo alle prese con un "utensile intelligente, capace di realizzare, a seconda del momento, intenzioni differenti", assumendo appunto una "certa autonomia, nella misura in cui il suo 'comportamento' non sarebbe dettato a ogni istante da un agente umano". I robot "intelligenti" inoltre sono in grado di "adattarsi da sé a una certa varietà di missioni" (Adler 2024, 177) L'autonomia è anche quella dell'apprendimento automatico, degli algoritmi "che possono imparare cose nuove interagendo col mondo. (…) Una IA deve avere la capacità di imparare cose nuove da sola, piuttosto che seguire le istruzioni dei suoi creatori originari"

(Harari 2024, 389-390). Questo appare quindi come un tratto distintivo rispetto alle tecnologie precedenti.

L'IA è quindi una "risorsa crescente di capacità di *agire* interattiva, autonoma e spesso autoapprendente, (…) una *riserva di capacità di agire a portata di mano*", che "*inscrive* il mondo" generando artefatti che interagiscono con il mondo e lo costruiscono (Floridi 2022, 53-54). Proprio in virtù della sua autonomia l'IA *scrive* il mondo, entra nel processo di descrizione del mondo perché contamina il linguaggio e le modalità di dialogo, fattori performativi. Se, come taluni sostengono, l'uomo diventa macchina proprio mentre la macchina tende a umanizzarsi, la riflessione sull'autonomia dei sistemi di IA diventa forse uno dei capitoli principali della riflessione sociologica applicata al lavoro. Un aspetto che, legato al cambiamento nel lavoro citato poco sopra, segna un passaggio *antropologico* che richiede anche una disamina sotto il profilo *etico*. L'IA genera decisioni in maniera autonoma rispetto alla nostra mente, è in grado di procedere piano piano a imporre sistemi di decisione che le sono propri e che riteniamo convenienti da adottare perché capaci di farci arrivare velocemente in fondo al processo operativo. L'IA *raccoglie* informazioni e le *finalizza*, ma – e questo è un altro oggetto di studio socio-lavorista per i prossimi anni – è l'intervento umano che le *interpreta*, così come interpreta i risultati ottenuti dai processi *gestiti* dall'IA, e quindi, in ultima istanza, pare riappropriarsi di quella chiusura del cerchio che normalmente si tende ad attribuire a una tecnologia che decide autonomamente. La capacità generativa dell'IA, in forma *autonoma*, entra direttamente nei cicli di valorizzazione del capitale, e ci richiede un dispiego di (nuovi e/o rinnovati) strumenti concettuali e operativi per interpretare i processi di generazione del valore e le nuove forme di alienazione della persona rispetto al lavoro. Entra direttamente nella produzione delle diverse forme di capitale, economico, culturale, sociale e simbolico; entra nella relazione fra produzione e riproduzione.

(c) Controllo

L'IA ripropone al centro dell'interesse dell'analisi socio-lavorista il tema del controllo. La pervasività dell'IA, che tende ad allontanarsi dal diretto controllo umano acquisendo sempre più autonomia, condizione perché possa essere ritenuta appunto *intelligente*, pone sul tavolo il problema del controllo da due punti di vista, dalle due direzioni che caratterizzano il (nuovo) rapporto uomo-macchina: controllo del lavoratore verso la macchina e della macchina verso il lavoratore. La macchina per conquistarsi la qualifica di *intelligente* deve sempre più allontanarsi dalle istruzioni dirette e continue dell'operatore, per adottare autonomamente comportamenti. Secondo alcuni autori, i sistemi di Intelligenza artificiale necessitano dell'intelligenza umana per strutturarsi, finendo per generare sistemi di controllo dei comportamenti umani. Le persone, nel lavoro e nella vita in generale, spesso senza soluzione di continuità, subiscono il controllo e il governo dell'Intelligenza artificiale, ma anche le soluzioni tecnologiche altamente sofisticate richiedono l'intervento umano, in quanto sono gli esseri umani "che portano a termine compiti che le macchine non possono svolgere" (Crawford 2021, 250). In ogni caso la preoccupazione di perdere il controllo è sempre presente: "una delle nostre paure nei confronti dell'Intelligenza artificiale ha a che fare con la possibilità che la nostra individualità rimanga schiacciata da algoritmi che decidono per noi. Anche se le macchine non volessero ridurci al silenzio, potrebbero volerlo fare coloro che le controllano: puoi pensarla così oppure no, ma il computer dice no, quindi taci" (Runciman 2024, 7-9).

Potremmo però anche in questo caso leggere il tema del controllo secondo la prospettiva dell'analisi della qualità del lavoro e della vita lavorativa (Gosetti 2022), quindi come possibilità per il lavoratore di condividere i processi decisionali che definiscono le condizioni lavorative, e, in primo luogo, gli aspetti di organizzazione e contenuto del lavoro. Si torna quindi al tema della partecipazione, nelle diverse declinazioni (individuale e collettiva, diretta e indiretta, informale e formale ecc.) e a differenti livelli (informazione, consultazione, co-decisione) come strategia per *generare controllo* da parte del lavoratore. L'IA può agire direttamente sulla gestione dei "margini di incertezza" (Crozier e Friedberg 1978), che è alla base delle relazioni nelle organizzazioni e costitutiva del potere organizzativo. I sostenitori della superiorità umana sulle macchine ci ricordano che "gli esseri umani sono la principale fonte dell'incertezza" e che "quando sono coinvolte delle persone, la fiducia in algoritmi complessi può creare illusioni di certezza" (Gigerenzer 2023, 13). Proprio su questi aspetti però si vanno (ri)giocando le relazioni nelle organizzazioni altamente tecnologizzate,

sull'asimmetria di controllo dei margini di incertezza organizzativa e, in ultima analisi, sulle possibilità che uno strumento potente come l'IA ha di consentire a chi governa le organizzazioni di favorire/comprimere l'autodeterminazione del lavoratore (Gosetti 2022). L'IA può essere uno strumento potente per creare e rafforzare asimmetrie nei luoghi di lavoro, e quindi favorire in diversi modi logiche di dominio.

(d) Generatività

Una distinzione che va operata è quella fra l'IA che si limita a riprodurre quello che ha immagazzinato, a riproporlo tale e quale o semmai elaborato, rispondendo a richieste di informazioni, e l'IA che genera qualcosa di nuovo, innesca produzioni di elementi non immagazzinati (tesi, immagini, musica ecc.). Il salto di qualità, secondo molti autori, è stato proprio quello di proporsi come IA di carattere *generativo* (Kaplan 2024) La generatività dell'IA applicata al lavoro diventa anche generatività in termini socio-tecnici. Quindi non solo nelle modalità dette sopra, ma in una nuova compenetrazione e socio-tecnica. Utile diventa quindi rileggere i mutamenti in atto nel lavoro nei termini che ci ha insegnato il Tavistock Institute (Gosetti 2024), osservando i processi di implementazione ed evoluzione delle tecnologie radicati entro contesti sociali di utilizzo, e come questo processo di compenetrazione debba costituire anche una postura nella progettazione della tecnologia. Dobbiamo cominciare a riflettere inoltre sulle possibilità che le macchine generative siano in grado anche di produrre autonomamente istruzioni alle quali adeguare il loro comportamento. Si tratta di riflettere quindi sulla relazione fra prevedibilità e imprevedibilità, probabilità e improbabilità dei comportamenti. La prospettiva da taluni paventata è che l'IA crei universi paralleli al nostro (Domingos 2018), che, affiancandoci, definisca percorsi di vita altamente prevedibili che prendano il sopravvento sulla nostra imprevedibilità e capacità di farci sorprendere dalle scoperte inaspettate, di essere, in definitiva, soggetti capaci di indeterminatezza.

Il carattere di generatività contribuisce a traslare l'IA verso l'intelligenza collettiva, produce elementi nuovi che influenzano insiemi collettivi di riflessione attorno all'elemento generato. Può favorire lo sviluppo di comunità e diventare quindi generatrice di comunità. La stessa informazione prodotta dall'IA diventa conoscenza nella misura in cui viene inserita in una struttura comunitaria, e quindi 'genera' comunità. Un tema di studio interessante è quindi quello del processo di contemporanea collettivizzazione e incorporazione dell'IA, della compenetrazione a livello micro (persona), meso (gruppi e organizzazioni) e macro (sistemi di produzione e riproduzione sociale) che sta alla base del carattere generativo dell'IA.

(e) Scelta

A partire da quanto sostenuto finora, formuliamo anche un'ulteriore ipotesi interpretativa: l'IA è una tecnologia che *decide*, ma non *sceglie*, prende decisioni, ma non si relaziona con universi valoriali a partire dai quali adotta alcuni comportamenti piuttosto che altri, *scegliendo* appunto in relazione a *opzioni di valore*. La *decisione* è un processo computazionale trasferibile su sistemi che trattano dati per risolvere problemi; la *scelta*, sebbene possa centrarsi anch'essa su sistemi di calcolo che prevedono di processare informazioni, richiede un *giudizio* ancorato a valori (Perilli 2025). Harari ci dice che "mentre perseguono vari obiettivi, i computer in rete sviluppano un modello comune del mondo che li aiuta a comunicare e cooperare" (Harari 2024, 396). Le tecnologie coordinandosi producono ordine, quindi generano processi di decisione che ordinano attraverso azioni, ma non sono processi di scelta, se per scelta intendiamo l'ancoraggio dell'azione, riflessivo, preventivo e in itinere, a universi valoriali, condivisi e condivisibili. L'IA chiama direttamente in causa la dimensione etica, quindi l'aggancio delle scelte a universi di valori, anche quando si tratta di progettare e adottare l'IA nel lavoro. Più in generale vi è chi ha parlato di "design etico", "*values in design*" (Adler 2024), quando siamo in presenza di una progettazione della tecnologia che discute le implicazioni per il lavoratore, l'essere umano al lavoro, considerando che "il problema centrale dell'etica dell'IA non è di ordine teorico, ma pratico" (Adler 2024, 359). Si tratta, infatti, di verificare costantemente in che modo nella concretezza dei meccanismi operativi quotidiani l'IA agisca sui limiti e le potenzialità del lavoratore (quindi sull'autodeterminazione) e sugli universi valoriali che sono posti a riferimento della discussione su di essi/e (limiti e potenzialità).

Come spesso viene ricordato l'IA può muoversi entro un universo di pregiudizi. L'addestramento delle reti neurali, operato da esseri umani con i

loro universi socio-culturali di riferimento (valori e valutazioni), può portare con sé elementi di pregiudizio, trascinare dentro i sistemi decisionali, in maniera più o meno consapevole, stereotipi consolidati in grado di produrre discriminazione. Problema che, secondo alcuni potremmo risolvere se "dall'addestramento delle macchine si escludesse del tutto ogni decisione umana" (Spitzer 2024, 230). Argomentazione di indubbio carattere provocatorio, che ha il vantaggio però di farci di nuovo riflettere sulla necessità di trasparenza e di controllo delle responsabilità umane nella progettazione e nell'addestramento delle macchine, così come nel loro utilizzo. Anche in questo caso abbiamo sul tavolo alcuni binomi che sollecitano ulteriormente le nostre riflessioni, anche guardando al futuro: determinismo vs libertà, predeterminazione vs responsabilità.

Le macchine adottando una razionalità strumentale, potente, orientata a risolvere problemi, secondo una modalità di ragionamento deduttivo, lineare, mentre l'intelligenza umana è capace di approcciare alla realtà anche secondo una logica *abduttiva*, che rivede continuamente le proprie decisioni lungo un percorso di verifiche progressive di ipotesi. A fare la differenza è forse proprio il fatto che nel processo decisionale la componente umana inserisce un coefficiente di scelta legato all'attribuzione di valore, condizione non contemplata dalla decisionalità della macchina. La decisione dell'IA chiude il processo, anche velocemente, raggiungendo l'obiettivo prefissato, e anche quando raccoglie elementi per modificare la sua azione, quindi è allenata ad apprendere, tende a codificare un comportamento in vista di un obiettivo, mentre la scelta dell'intelligenza umana apre, asseconda l'imprevedibilità, la pone a confronto con universi valoriali di riferimento.

(f) Habitus

Mentre cerca di avvicinarsi il più possibile al linguaggio umano per interagire con l'essere umano, l'IA interviene sul linguaggio, modificandolo anche profondamente. La relazione lavoratore-macchina ridefinisce il linguaggio, creare nuovi linguaggi, e il linguaggio, come sappiamo da ormai tanto tempo, è uno dei principali fattori per generare valore (economico). L'IA agisce su alcuni riferimenti socio-culturali della nostra interazione umana, modifica prassi relazionali, costruisce mondo, trasforma modalità espressive, e nel fare questo agisce sui riferimenti normativi della relazione interpersonale, entra nella relazione fra attore e sistema e nei processi a doppia strutturazione di cui spesso ci parla la sociologia. Noi strutturiamo il mondo con le nostre rel-azioni, ma agiamo dentro un mondo strutturato che (im)pone riferimenti (materiali e culturali) alle nostre rel-azioni, e così via. E l'IA entra potentemente in questo processo di continua *doppia strutturazione* sociale strutturata. Le nuove forme di generazione del valore economico che vedono l'IA attraversare immaterialità e corporeità, essere sempre più indossata, entrano a far parte dell'habitus. Il valore generato dall'IA non ha bisogno di un luogo di lavoro preciso. Si tratta di un'immaterialità deterritorializzata, che prevede però 'luoghi' di controllo, potenziali e reali situazioni di dominio legate alla capacità di interagire con l'IA, rendendola interlocutore. Allo stesso tempo l'IA agisce sul corpo, viene incorporata, entra nella sfera emotiva *(emotion AI*), capta sentimenti, motivazioni, atteggiamenti, usa dati per studiare le emozioni e i comportamenti. Legge le intonazioni della voce, le espressioni del volto, il linguaggio del corpo. Di fatto entra direttamente in relazione con quello che Bourdieu chiamava l'habitus, l'insieme di disposizioni durevoli e orientamenti maturati nel corso della vita, che sono alla base delle nostre azioni (intenzioni, aspirazioni ecc.). Mettendo assieme dati oggettivi, caratteristiche fisiche, dati comportamentali, informazioni sugli atteggiamenti, espressioni linguistiche e facciali, entra nel nostro habitus, e quindi nei nostri universi di costruzione del senso (McQuaid 2022). Dell'Intelligenza artificiale dobbiamo cogliere pertanto anche il suo aspetto materiale, *incarnato*, osservando come metta assieme risorse naturali, lavoro umano, storie personali e collettive, classificazioni (Crawford 2021). Potremmo dire che finisce per essere incorporata in un habitus, costrutto di disposizione immateriali e materiali, di identità e vitalità alla base del nostro agire. Diventa *generativa* anche nei confronti del nostro habitus.

Un capitolo particolarmente interessante sul quale cimentare l'attenzione della ricerca sociologica dei prossimi anni è quello della diffusione dei cosiddetti 'nanorobot', componenti elettroniche di ridottissime dimensioni, che potremmo *incorporare* per diverse finalità e che di fatto agiranno sul nostro habitus, sulle nostre modalità di vita e di presa delle

decisioni. Componenti che tenderanno a generare un costrutto in cui soluzioni tecnologiche e corpo umano fisico si fondono senza soluzione di continuità.

(g) Voice

Pur non avendo *coscienza* di quello che genera e del processo attraverso il quale genera, l'IA ha *conoscenza*, e per certi versi conserva memoria, di quello che genera, perché controlla le sue azioni, è attenta ai – e valuta i – risultati di esse. È una tecnologia che nel lavoro quotidiano, a partire dai dati che raccoglie sul comportamento adottato, è capace di riposizionarsi nei processi operativi e nel contesto più complessivo. Questo aspetto ci porta alla necessità di comprendere quali responsabilità vengono attribuite ai sistemi di IA, e come si generano responsabilità e si attribuiscono responsabilità ai vari ruoli implicati relativamente alla progettazione e al funzionamento dei sistemi di IA. Secondo alcuni l'IA di fatto ha una capacità di risolvere i problemi che non richiedono un 'libero arbitrio' e sostanzialmente rappresenterebbe un'estensione delle nostre capacità (Domingos 2018). Ma anche questi sono interrogativi sui quali dovremo riflettere nei prossimi anni. Infatti, l'IA ci riporta all'esigenza di creare contesti di confronto e di scelta condivisi. Pensare e progettare *luoghi* del lavoro, anche quando fisicamente le attività non richiedono una localizzazione precisa e visibile, che prevedano possibilità di *voice*, di agire criticamente verso le soluzioni che riguardano l'IA applicata a modelli organizzativi e contenuti del lavoro. Non devono quindi mancare campi, perimetri spazio-temporali, che in qualche modo possano rappresentare territori di senso, all'interno dei quali condividere le scelte relativamente a ragioni (perché) e modalità (come) di adottare le tecnologie dell'IA. Potremmo pensare ad *arene* dentro le quali si possa ritrovare quell'agire comunicativo, per certi versi habermasiano, che vede soggetti nella condizione di poter *discutere* l'adozione e le implicazioni dell'IA, controllando i potenziali effetti distorsivi della comunicazione. Affinché le parti in causa possano interpretare un diritto di *voice*, quindi, condividere la progettazione dell'organizzazione e dei contenuti del lavoro in presenza dell'IA, devono comunque essere preparate per farlo. Per agire una critica sono necessarie competenze e spazi nelle quali esercitarla. Già poco sopra, a questo proposito, abbiamo fatto cenno alla partecipazione e alle diverse declinazioni nelle quali si concretizza nei luoghi di lavoro. L'esercizio del diritto di *voice*, richiede anche possibilità di verifica dell'esito delle istanze sottoposte a discussione, argomentate nelle *arene di discussione* dell'organizzazione (assemblee, riunioni, gruppi di lavoro ecc.). Quindi si tratta di una facoltà di agire nei contesti di giustificazione che discutono e legittimano le scelte organizzative.

Un aspetto che va studiato è se la stessa IA può costituire anche uno strumento di voice, sia per coinvolgere nella raccolta di suggerimenti, creando senso di appartenenza, sia per far emergere istanze, quindi per la voce alle persone dentro i luoghi di lavoro. Il pericolo di cadere in un 'algocrazia', un sistema di lavoro che vede gli algoritmi al governo, può essere scongiurato aumentando la conoscenza dei sistemi e la capacità di lettura critica, per non vedere la possibilità di governo, individuale e collettiva, espropriata dalla capacità tecnologica (tecnocrazia). Nelle agorà si dovrà discutere di etica e IA, del legame che intercorre fra innovazione, tecnologia e democrazia, mettendo a tema l'influenza che l'IA è in grado di esercitare sui processi sociali, sui comportamenti, ma ancor più sulle intenzioni e la maturazione delle convinzioni delle persone (sull'habitus citato poco sopra). Questo è un territorio nel quale ripensare anche le forme e modalità della rappresentanza collettiva, nel quale ridisegnare le modalità di relazione fra le parti e di contrattazione collettiva.

Pensiamo si possa chiudere questa prima parte mettendo al centro delle nostre riflessioni l'ipotesi che l'IA possa finire per rafforzare quella clessidra che nel (mercato del) lavoro abbiamo visto generarsi negli ultimi anni, vale a dire la polarizzazione (e quindi il distanziamento) fra una parte alta e una bassa del (mercato del) lavoro, soprattutto in termini di qualità del lavoro e della vita lavorativa. Un'ipotesi che già da prima della digitalizzazione interessa le analisi socio-lavoriste, e che anche l'avvento dell'IA pare vada confermando. A questo proposito, dobbiamo primariamente fare riferimento al "lato oscuro dell'IA", a quella sfera invisibile di lavoratori che generano dati, "lavoratori sottopagati necessari per contribuire a costruire, mantenere e testare i sistemi di Intelligenza artificiale", "microlavoratori" dediti a compiti digitali molto ripetitivi (etichettatura e caricamento di dati, revisione di contenuti ecc.) (Crawford

2021, 75). Una serie di attività a basso costo, non eseguibili dalle macchine o più convenientemente realizzabili da esseri umani, che genera un'area di sfruttamento lavorativo e alienazione. Ma l'IA, come abbiamo detto poco sopra, è anche uno strumento che potenzia significativamente le capacità del lavoratore, crea un lavoratore potenziato, in grado di sviluppare azioni 'abilitate' proprio dalla tecnologia di ultima generazione. Una tecnologia che crea nuovi lavori non necessariamente di bassa qualità, anzi. Ma dovremmo produrre ricerca proprio per individuare ipotesi analitiche da sottoporre a verifica, per cogliere le possibilità e modalità di partecipazione ai processi di progettazione delle tecnologie e quindi della qualità della vita lavorativa. Non siamo in grado di dire come si strutturi questa nuova fase di polarizzazione, come si vadano generando o consolidando asimmetrie fra datore di lavoro e lavoratore, come si possano ridurre le disuguaglianze, di genere e generazionali, in primo luogo legate alla richiesta di nuove competenze. Di fatto ci aspettano tante tappe di lavoro analitico sul campo e di confronto interdisciplinare.

Due su tutte ci paiono comunque le piste di lavoro prioritarie sulle quali avviare un confronto: (a) il tema dei valori e (b) i riflessi dell'introduzione dell'IA sulle dinamiche collettive nei luoghi di lavoro.

Già nelle pagine precedenti ci siamo soffermati sul tema dei valori. Lo abbiamo fatto nel sostenere che la macchina intelligente decide e non sceglie, quindi non ancora la propria azione a un universo valoriale di riferimento, che orienta appunto la scelta. Produce una decisione rispettando una prescrizione, un format procedurale al quale si attiene. Ma qualcuno potrebbe obiettare, giustamente, che la macchina, quando viene *addestrata a decidere*, assume i *valori dell'addestratore*. E il caso diventa ancora più intrigante sotto il profilo analitico quando l'addestramento lascia aperta alla macchina la *possibilità di apprendere in corso d'opera*. Se ipotizziamo per un momento che, come ci suggerisce Dejours, lavorare significhi "*colmare lo scarto* tra il prescritto e l'effettivo", e che "ciò che bisogna mettere in atto per colmare questo scarto non [possa] essere previsto in anticipo", è chiaro che il lavoratore deve mettere in campo competenze (e intelligenza) per attraversare il "cammino (…) tra il prescritto e l'effettivo" (Dejours 2020, 8). Questo può implicare anche mettere sotto una lente critica le prassi operative, porre in discussione l'addestramento e l'addestratore della macchina. In ultima analisi, assumere la *critica* come valore di riferimento. Potremmo anche supporre che l'incertezza, l'instabilità dei problemi, apra spazi che non possono essere occupati dall'apprendimento e agire automatico dell'IA (Gigerenzer 2023). O, da un'altra prospettiva, che partendo proprio dalla capacità dell'IA di generare decisioni certe, traiamo spunto per pensare a scelte instabili, a produrre discontinuità di percorso e di risultato. E qui di nuovo si apre il confronto sui valori, come orientatori e mediatori sociali alla base dei processi di condivisione delle scelte di discontinuità. Un confronto che deve prendere le mosse primariamente da una riflessione sugli effetti delle decisioni che l'IA genera nel lavoro (per stare al contesto che qui stiamo privilegiando). L'attenzione della ricerca può quindi essere posta sulle diverse forme che assume il lavoro, dentro *luoghi* tradizionali o in nuove *prassi* di generazione dei risultati. Quindi, se anche le macchine agiscono a partire da valori, potremmo dire, incorporati, si tratta di capire quanto i lavoratori siano in grado di (vi siano condizioni e competenze per) agire il valore della discontinuità, dell'indeterminatezza (a livello individuale e/o collettivo). Dato che le analisi ci dicono che l'IA sta generando e rafforzando asimmetrie nei luoghi di lavoro, e quindi le decisioni prese dalle macchine intelligenti vanno a vantaggio soprattutto di alcune parti in gioco, la prima discontinuità da mettere sul tavolo della discussione potrebbe essere legata ai valori dell'uguaglianza e della libertà nel lavoro. Si tratta quindi di analizzare i processi di regolazione dell'IA e di regolazione delle applicazioni dell'IA, per comprendere come il lavoro sia un contesto di libertà e generatore di libertà.

Altro grande tema, che per certi verso si lega a quello appena argomentato, riguarda la facoltà intrinseca delle recenti innovazioni tecnologiche di agire sulla connessione, come lo sono state da sempre quelle digitali, che nella versione dell'IA paiono però in grado di farlo con un elevato livello di autonomia organizzativa. Una connessione non solo fra tecnologie e fattori della produzione, ma anche fra persone, fra individui che cercano senso attraverso (e nell') l'*agire individuale e collettivo*. L'*agire collettivo* dentro l'organizzazione si lega direttamente alla dimensione del *riconoscimento* come concetto chiave di una lettura sociologica del lavoro. Il riconoscimento è un concetto relazionale,

come possiamo apprendere da tanta letteratura sociologica a partire dagli studi di Axel Honneth. E recentemente proprio Honneth ci invita a riflettere sul ruolo del lavoro come risorsa per "contribuire in libertà, senza coazione e senza vergogna, alla prassi della formazione democratica della volontà" (Honneth 2025, 8). Un secondo filone di studio potrebbe quindi mettere a tema quanto l'IA agisca direttamente sulla formazione dell'agire collettivo dentro l'organizzazione, quanto sia in grado di generare modelli di comportamento, lasciando aperti o chiudendo spazi per il lavoro di squadra e l'azione collettiva più in generale. Significativi processi di individualizzazione in atto ormai da anni nel lavoro, associati a nuove frammentazioni delle attività, hanno indebolito le dimensioni collettive, producendo spesso isolamento del lavoratore. Un isolamento favorito anche da precise scelte organizzative, prese anche senza il diretto coinvolgimento del lavoratore o delle sue rappresentanze. L'IA è una tecnologia di organizzazione perché è una tecnologia di comunicazione, che usa un linguaggio e genera un linguaggio. E proprio la generazione di un linguaggio (non solo nelle organizzazioni lavorative, ovviamente) crea comunità di appartenenza e condizioni di esclusione. Se mettiamo sul tavolo varie ipotesi che stanno circolando relativamente alle nuove tecnologie; (i) che agiscono direttamente sul 'senso del noi', definiscono specificità delle comunità digitali (che condividono obiettivi e interessi) rispetto a quelle naturali (che condividono spazi e tempi) (Riva 2025), ridisegnando la tradizionale contrapposizione fra comunità e società ereditata dalla sociologia di Tönnies (1887); (ii) che generano strutture comunicative (pattern) (Esposito 2022), manipolando dati, generando percezioni della realtà, centralizzando le risorse necessarie a produrre influenza (potere) e nuove forme di controllo; che propongono nuove modalità operative in grado di agire sui concetti di intelligenza e di coscienza così come finora li abbiamo intesi (Perilli 2025); e così via.

Se tutte queste sono traiettorie di riflessione imprescindibili per la ricerca sociologica, siamo allora chiamati a studiare e comprendere come le tecnologie dell'IA ri-definiscano modelli di comportamento collettivi, che diventano confini alle scelte individuali. Le tecnologie di ultima generazione, compresa l'IA alla quale ci riferiamo in questo saggio, attraversano dimensioni individuali e collettive. La sociologia del lavoro ha dedicato negli anni molti studi al tema della partecipazione e alle condizioni che la rendono effettiva. Come abbiamo evidenziato, l'IA, agendo direttamente sui processi decisionali, investe le condizioni che producono una partecipazione realmente incisiva, a livello individuale e collettivo. È una tecnologia che dobbiamo *problematizzare*, nel senso di tradurre in grandi argomenti sfidanti per la ricerca sul campo, proprio mettendo a tema principalmente il coinvolgimento del lavoratore, da solo e in forma associata. Se da anni ormai le analisi, divenute letteratura consolidata e acquisita per la disciplina, ci parlano di una individualizzazione dei processi lavorativi, con relativo disancoramento del lavoratore da universi di solidarietà più o meno istituzionalizzati che garantivano tutele e diritti (Castel 2004 e 2019; Sennett 2000), dobbiamo a questo punto chiederci quali effetti provocherà l'IA proprio sul coinvolgimento della persona al lavoro e nel corso della sua vita lavorativa. A tema di ricerca finiscono pertanto sia i modelli organizzativi e quindi le modalità di collaborazione a livello verticale e orizzontale che si stanno (ri)disegnando nel lavoro, sia le forme di rappresentanza che devono necessariamente essere ripensate per intercettare la forte eterogeneità che anche le innovazioni tecnologiche contribuiscono a generare nel lavoro e nella vita lavorativa.

2. Intelligenza artificiale e tutele sul lavoro: l'impatto dell'AI Act nel sistema integrato delle fonti

La (ri)costruzione delle regole di tutela del lavoratore di fronte all'uso datoriale di sistemi di IA richiede di verificare le fonti normative applicabili nell'ambito di processi organizzativi in cui possono essere veicolati, nascosti e/o amplificati poteri datoriali direttivi e di controllo, trattati dati personali, introdotte e/o esacerbate fonti di pericolo per la salute e sicurezza, la dignità, la libertà, nonché per il diritto del lavoratore a non essere discriminato (*ex multis*, Paliokaité *et. al.* 2025; Lai 2025; Treu 2024; Novella 2022; Tullini 2021; Tebano 2020). La sfida regolatoria non si muove, tuttavia, solo sul piano della mappatura dei raccordi e delle implicazioni normative: da un lato, si pone il confronto con fonti che non presentano una matrice lavoristica, come ad esempio il Regolamento sulla protezione dei dati

personali (GDPR), dall'altro, vi è la necessità di non perdere l'orizzonte della proiezione collettiva delle tutele, espressione chiave dell'impronta propria del diritto del lavoro, collegata all'applicazione del principio di eguaglianza sostanziale. Su questa sfida si inserisce ora l'impatto dell'AI Act, normativa di prodotto chiamata a tradurre il complesso equilibrio tra protezione dei diritti fondamentali e sviluppo della competitività di mercato, tra la scelta di un modello regolativo *hard* e di dettaglio e la rapida evoluzione che connota la categoria tecnologica in questione. L'indagine che segue cerca di analizzare la risposta regolativa UE sull'IA e la sua rilevanza per l'ambito del lavoro, partendo da alcuni punti di osservazione: il suo inquadramento e la sua tenuta nel contesto internazionale, i termini con cui traccia le fattispecie di riferimento, anzitutto quella di 'sistema di IÀ, e articola l'approccio basato sul rischio, il carattere minimo e complementare del sistema di regole predisposto e il conseguente ruolo delle Parti sociali.

Confini di regole…

Nella relazione di accompagnamento alla proposta di Regolamento sull'IA, la Commissione europea affermava che l'interesse dell'Unione è "preservare la leadership tecnologica dell'UE", nonché "tutelare la sovranità digitale dell'Unione e sfruttare gli strumenti e i poteri di regolamentazione di quest'ultima per plasmare regole e norme di portata globale" (par. 1.1. e par. 2.2)[2].

Come è stato sottolineato dalla dottrina, tali dichiarazioni di intenti si confrontavano con un dato di fatto: l'Europa non è un produttore di tecnologia, né possiede importanti piattaforme o sistemi di comunicazione digitale, a differenza di Cina e Stati Uniti. Il piano su cui poteva o può eventualmente competere è quello della regolamentazione, attraverso l'innesco del c.d. 'effetto Bruxelles', già ottenuto con il GDPR (Bradford 2020)[3], con conseguente affermazione del modello normativo UE di risposta all'IA come punto di riferimento a livello globale (Finocchiaro 2024). In tale ottica, si ponevano e si pongono le norme sull'applicazione extraterritoriale contenute nell'AI Act, volte a vincolare alle garanzie UE anche fornitori e deployer stabiliti in Paesi Terzi quando il sistema o modello sia immesso nel mercato europeo, ovvero quando il sistema produca un output utilizzato nell'Unione (v. art. 2, par. 1, Reg. 2024/1689; Bassini 2024).

A pochi giorni dall'entrata in applicazione delle prime disposizioni dell'AI Act, l'impressione data da alcune scelte della Commissione europea è che a livello globale si stia registrando non tanto un 'effetto Bruxelles' quanto un 'effetto *su* Bruxelles'.

L'inquadramento delle fonti UE nel più ampio contesto internazionale aveva permesso fino ad ora di rilevare come i diversi livelli di azione condividessero alcune scelte di fondo, metodologiche e di contenuto, in particolare un approccio basato sul rischio e l'individuazione, quali garanzie imprescindibili di una prospettiva antropocentrica, della trasparenza e supervisione umana (Peruzzi 2023; Iannuzzi 2025).

In tale prospettiva si sono collocati, in particolare, la Convenzione del Consiglio d'Europa (di cui fanno parte tutti i 27 Stati membri dell'UE), completata a marzo 2024 e firmata il 5 settembre 2024 (tra gli altri) dall'UE, gli Stati Uniti, il Regno Unito e Israele[4], e il Processo di Hiroshima. Quest'ultimo, condotto dai leader del G7, ha portato, prima, il 30 ottobre 2023, all'adozione di principi-guida internazionali e di un Codice di condotta internazionale per le organizzazioni che sviluppano AI avanzata, poi, il 13 settembre 2024, alla sottoscrizione di un Piano d'azione, in cui si valorizzano, tra i vari profili, da un

2 Si veda *Proposta di Regolamento che stabilisce regole armonizzate sull'intelligenza artificiale*, COM(2021) 206 final.

3 Rispetto all'applicazione extraterritoriale del Regolamento (UE) 2016/679 sulla protezione dei dati personali, si segnala la recente vicenda che ha interessato l'applicazione cinese *DeepSeek* e il blocco della stessa da parte di molti garanti privacy europei a fronte dell'assenza di garanzie fornite circa la conformità del trattamento dei dati personali degli utenti dell'UE alla normativa europea in materia. Come rileva il Provvedimento n. 33 del 30 gennaio 2025 dell'authority italiana, il gestore del servizio era vincolato al rispetto del GDPR in forza di quanto stabilito dall'art. 3, par. 2, secondo cui il "regolamento si applica al trattamento dei dati personali di interessati che si trovano nell'Unione, effettuato da un titolare del trattamento o da un responsabile del trattamento che non è stabilito nell'Unione, quando le attività di trattamento riguardano: a) l'offerta di beni o la prestazione di servizi ai suddetti interessati nell'Unione".

4 Convenzione quadro sull'Intelligenza artificiale, i diritti umani, la democrazia e lo stato di diritto. Le precondizioni poste per la firma da alcuni Paesi, in particolare dagli Stati Uniti, spiegano il ristretto ambito di applicazione della Convenzione, costituito dalle attività del ciclo di vita dei sistemi di Intelligenza artificiale intraprese dalle autorità pubbliche o da attori privati che agiscono per loro conto (art. 3, par. 1, lett. a).

lato, il principio di accountability, in base al quale "la responsabilità per le decisioni concernenti i rapporti di lavoro, anche qualora prese da sistemi di IA, dovrebbe rimanere in capo ai datori di lavoro", dall'altro, l'importanza del dialogo sociale e della contrattazione collettiva a tutti i livelli[5].

La prospettiva di una IA affidabile, sicura e priva di discriminazioni, e di uno sviluppo responsabile, basato su tutela della trasparenza e garanzia del controllo umano, erano principi che, inoltre, si ponevano alla base dell'Executive Order adottato il 30 ottobre 2023 dalla Presidenza Biden[6].

Il quadro globale è, tuttavia, radicalmente mutato, con la revoca di quest'ultimo ordine esecutivo da parte di Trump il giorno stesso dell'insediamento alla Casa Bianca e la mancata firma di Stati Uniti e Regno Unito alla dichiarazione finale dell'AI Action Summit di Parigi dell'11 febbraio 2025 (firmata invece dalla Cina)[7]. Sempre al Summit di Parigi J.D. Vance, Vicepresidente degli Stati Uniti, ha criticato aspramente l'approccio a suo dire iper-regolativo dell'Unione: poche ore dopo, la Commissione europea ha ritirato la proposta di Direttiva in materia di responsabilità civile per danni da IA[8], che pur era rimasta ferma da due anni, non senza critiche da parte di molti giuristi. A smentire il collegamento tra le critiche e il ritiro è stata la Commissaria europea per la sovranità tecnologica Henna Virkkunen in una intervista al *Financial Times* del 14 febbraio 2025.

Sempre l'11 febbraio, dando seguito al documento strategico *Bussola per la competitività dell'UE* adottato a gennaio, la Commissione europea ha pubblicato il Programma di lavoro 2025, fortemente incentrato sulla semplificazione regolativa, come sottolineato sul sito istituzionale. In tale sede, oltre al primo intervento c.d. *omnibus* in materia di sostenibilità, incidente, tra le altre, sulla neo-adottata Direttiva *Due diligence* 2024/1760, è annunciata una valutazione d'impatto sul pacchetto digitale, comprensivo di GDPR e AI Act. La strategia si pone in linea con il Report di Mario Draghi *The future of European competitiveness. A competitiveness strategy for Europe* del settembre 2024 in cui, con specifico riferimento all'IA, si segnalano le lodevoli ambizioni del GDPR e del nuovo Regolamento, ma anche "la loro complessità e il rischio di sovrapposizioni e discrepanze", tali da poter "compromettere gli sviluppi nel campo dell'IA da parte degli attori industriali dell'UE. (…) Poiché nella competizione globale sull'IA stanno già prevalendo le dinamiche 'chi vince prende di più', l'UE si trova ora ad affrontare un inevitabile compromesso tra tutele normative ex ante più forti per i diritti fondamentali e la sicurezza dei prodotti, e regole più leggere per promuovere gli investimenti e l'innovazione dell'UE, ad esempio attraverso il *sandboxing*, senza abbassare gli standard dei consumatori" (Report Part B, p. 79)[9].

Se la posizione della Commissione, articolata nel Programma di lavoro, è quella di riverificare o rivedere il punto di equilibrio tra la tutela dei diritti fondamentali e le esigenze di sviluppo tecnologico e relativa competitività di mercato, il tema di un possibile intervento normativo dedicato al lavoro continua, cionondimeno, a riproporsi nel dibattito istituzionale. La prospettiva, già promossa dall'European Trade Union Confederation (ETUC) in una risoluzione del 6 dicembre 2022, era stata condivisa sia da Gualmini e Benifei all'interno del Parlamento europeo, sia da Schmit, Commissario europeo per il lavoro e i diritti sociali nella Commissione von der Leyen I. Roxana Minzatu, attuale Vicepresidente esecutiva della Commissione europea per i diritti sociali, nelle osservazioni di apertura alla riunione della Commissione interparlamentare sull'IA e il mercato del lavoro del 17 febbraio 2025 ha confermato la necessità di una riflessione su un possibile ampliamento delle tutele, a partire dal modello fornito dal capitolo sulla

5 Ministerial Declaration, G7 Labour and Employment Ministers'Meeting in Cagliari 12-13 September 2024, *Towards, an inclusive human-centered approach for new challenges in the world of work*; Annex 1, *G7 Action Plan for a human-centered development and use of safe, secure and trustworthy AI in the World of Work*, spec. pp.16-18.

6 Executive Order 14110 of October 30, 2023, *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, Federal Register/Vol. 88, No. 210/Wednesday, November 1, 2023, p. 75191.

7 *Statement on Inclusive and Sustainable Artificial Intelligence for People and the Planet*, Palais de l'Élysée, Tuesday February 11th 2025.

8 Commissione europea, *Proposta di Direttiva del Parlamento europeo e del Consiglio relativa all'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'Intelligenza Artificiale (direttiva sulla responsabilità da Intelligenza Artificiale)*, Bruxelles, 28 settembre 2022, COM(2022) 496 final.

9 Si veda https://commission.europa.eu/topics/eu-competitiveness/draghi-report_en#paragraph_47059.

gestione algoritmica della Direttiva (UE) 2024/2831 in materia di lavoro mediante piattaforme digitali. Questa prospettiva ha trovato peraltro una prima declinazione nella proposta presentata dalla Commissione per l'occupazione e gli affari sociali del Parlamento europeo il 12 giugno 2025[10]. A ogni modo, il recepimento della direttiva piattaforme da parte degli stati membri, da effettuare entro il 2 dicembre 2026, costituirà un importante banco di prova per una verifica dei modelli normativi che meglio possono declinare nella prospettiva lavoristica le garanzie di trasparenza e sorveglianza umana rispetto all'utilizzo dei sistemi automatizzati, anche di IA.

... confini di fattispecie

Ragionare di tutele del lavoratore di fronte all'uso datoriale di sistemi di IA significa ragionare di sistema integrato di fonti regolative. È un'operazione a cui invita lo stesso AI Act nel momento in cui si qualifica come quadro normativo volto a stabilire uno standard di tutele minimo e complementare: minimo perché non osta all'introduzione di regole più favorevoli per i lavoratori a livello nazionale anche mediante contratto collettivo (art. 2, par. 11); complementare perché non deve pregiudicare le altre fonti UE o interne applicabili, essendo anzi chiamato a svolgere un ruolo funzionale, facilitante e abilitante nei confronti dei diritti e dei mezzi di ricorso esistenti (Cons. n. 9). Ciò trova esplicita conferma e declinazione nella disciplina del *deployer*, ossia di chi utilizza il sistema di IA sotto la propria autorità (art. 3, punto 4), soggetto centrale in una prospettiva d'analisi lavoristica, considerato che tale sarà, in questo quadro di regole, la (prima, ma non sempre unica, v. *infra*) qualificazione assunta da un datore di lavoro che faccia uso di sistemi di IA. Rispetto a tale figura, si chiarisce, in particolare, che l'obbligo di utilizzare il sistema seguendo le istruzioni consegnate dal fornitore non deve compromettere l'adempimento degli obblighi stabiliti da altre fonti (art. 26, par. 3). Nella (ri)costruzione dell'architettura regolativa applicabile, il raccordo è anzitutto con la disciplina del GDPR, se il sistema automatizzato implica il trattamento di dati personali, nonché, a livello di normativa nazionale, con l'art. 4 St. Lav. se lo strumento consente di raccogliere informazioni sull'attività lavorativa dei dipendenti, nonché con l'art. 1-bis D.Lgs. 152/1997 e gli obblighi informativi ivi previsti laddove si tratti di sistema decisionale o di monitoraggio integralmente automatizzato. Altrettanto fondamentale è il raccordo con la normativa in materia di salute e sicurezza e con quella antidiscriminatoria, entrambe di matrice UE. L'orizzonte della dimensione collettiva propria dell'ottica lavoristica invita, infine, a declinare questi profili anche nella prospettiva del ruolo delle Parti sociali, nei varchi d'accesso a loro forniti dalle diverse fonti coinvolte, ma anche negli interventi che a livello di contrattazione collettiva possono sviluppare, soprattutto a sostegno di un coinvolgimento delle rappresentanze dei lavoratori nei processi di valutazione e gestione del rischio e dell'impatto dell'IA.

L'operazione richiede, d'altra parte, a monte, di individuare i confini delle fattispecie regolate dall'AI Act, per comprenderne l'ambito di applicazione materiale e l'incidenza sul quadro normativo delle tutele.

Si è già potuto evidenziare nei paragrafi precedenti gli elementi costitutivi della definizione di sistema di IA prevista dall'art. 3, n. 1 (v. *supra*).

Alla luce degli orientamenti elaborati dalla Commissione europea a norma dell'art. 96, preme qui segnalare, anzitutto, come l'elemento dell'autonomia, qualificato dalla presenza di un certo grado di indipendenza dall'intervento umano, escluda dalla definizione i sistemi progettati per eseguire tutti i compiti attraverso funzioni azionate manualmente. Nel chiarire che l'autonomia non deve essere necessariamente completa, la Commissione spiega come a rilevare come dato-chiave sia la capacità del sistema di generare da solo l'output con cui interagisce con l'ambiente esterno, potendo l'input, in ipotesi, essere anche inserito manualmente. Al riguardo, è importante sottolineare – per non cadere in un'erronea sovrapposizione tra autonomia e apprendimento automatico, da un lato, autonomia e processo decisionale integralmente automatizzato, d'altro – che la suddetta capacità può

10 Progetto di Relazione recante raccomandazioni alla Commissione sulla digitalizzazione, l'Intelligenza artificiale e la gestione algoritmica sul luogo di lavoro – plasmare il futuro del lavoro (2025/2080(INL)). Relatore: Andrzej Buła. Le raccomandazioni alla Commissione indicano, in particolare, il contenuto di una proposta di direttiva sulla gestione algoritmica sul luogo di lavoro, con ambito di applicazione esteso a tutti i lavoratori e datori di lavoro nell'Unione, nonché ai lavoratori autonomi individuali e ai relativi committenti di servizi.

basarsi sull'applicazione di approcci sia di *machine learning*, sia *logic-based* e l'output può consistere non solo in una decisione (e in tal caso avremo in effetti un sistema decisionale integralmente automatizzato ai sensi dell'art. 22 GDPR e dell'art. 1-bis D.Lgs. 192/1997[11]), ma anche in una previsione (stima di valori sconosciuti a partire da valori noti), in contenuti (nel caso di sistemi di IA generativi) o in mere raccomandazioni (non applicate automaticamente). A confermare che un sistema di IA, anche ad alto rischio, possa solo sostenere – e non prendere – le decisioni sono le disposizioni dell'AI Act che obbligano il *deployer* a determinate garanzie di trasparenza laddove il sistema adotti o *assista* nell'adozione di decisioni (art. 26, par. 11), obblighi di informazione che devono includere l'indicazione del diritto della persona interessata, oggetto di una decisione adottata dal *deployer* sulla base dell'output della macchina, alla spiegazione dei singoli processi decisionali (art. 86; v. Cons. n. 93)[12]. Si deve, d'altra parte, sottolineare che l'assenza di un impatto sostanziale sull'esito del processo decisionale, sia esso umano o automatizzato, permette di escludere a norma dell'art. 6, par. 3, la qualificazione ad alto rischio dei sistemi di IA elencati nell'allegato III, tra cui quelli utilizzati per la selezione e gestione del personale, salvo non effettuino profilazione di persone fisiche. Alla luce di quanto esposto, si comprende perché la fattispecie "sistema di IA" non possa essere completamente sussunta nella categoria normativa di "sistema decisionale o di monitoraggio integralmente automatizzato"[13], prevista ad esempio nell'art. 1-bis del D.lgs. 152/97. In tal senso si rileva una criticità nella formulazione dell'art. 11 del D.D.L. governativo in materia di IA, approvato in via definitiva dal Senato il 17 settembre 2025, laddove sovrapponendo le due fattispecie (e dimenticando, peraltro, le rappresentanze sindacali)

11 Come spiegato da Trib. Palermo, 20 giugno 2023, sono da considerarsi sistemi di monitoraggio/decisionali integralmente automatizzati, ai sensi dell'art. 1-bis, quelli "che non prevedono l'intervento umano contemporaneo o successivo rispetto alla decisione del sistema o al monitoraggio dei lavoratori, che, poi, di norma, è finalizzato proprio all'assunzione da parte del sistema di una decisione che li riguarda. Sono, quindi, in quest'ambito, integralmente automatizzati i sistemi che non prevedono l'intervento umano nella fase finale della decisione o del monitoraggio, a prescindere da un eventuale intervento dell'uomo nelle fasi antecedenti, quale quella di mero inserimento di dati, comunque elaborati". In tal senso anche Trib. Torino, 5 agosto 2023: "un sistema decisionale integralmente automatizzato è quello che adotta decisioni senza richiedere l'intervento umano; (…) la circostanza che un intervento umano si possa rendere necessario per ovviare a problematiche che non attengono ad un percorso fisiologico, denota come l'intervento umano sia meramente eventuale, accessorio e, dunque, irrilevante". Rispetto all'art. 22 GDPR, si può richiamare la pronuncia della Corte di giustizia dell'Unione europea (CGUE), 7 dicembre 2023, C634/21, *Schufa*, ECLI:EU:C:2023:957, in cui si spiega che la nozione di 'decisione' a cui rinvia l'articolo non è definita dal Regolamento; a fronte della formulazione della disposizione, la stessa deve riferirsi "non solo ad atti che producono effetti giuridici riguardanti il soggetto di cui trattasi, ma anche ad atti che incidono significativamente su di esso in modo analogo". In tal senso, un determinato outcome, quale ad esempio un punteggio di solvibilità, rientra nel concetto di "decisione completamente automatizzata", qualora incida in modo decisivo sulle scelte assunte, in ipotesi anche da un terzo, in relazione a un rapporto contrattuale. Nella più recente pronuncia del 27 febbraio 2025 sul caso *CK* (C-203/22), la Corte ha altresì chiarito che Il diritto di informazione relativo ai processi decisionali automatizzati previsto dal GDPR copre "qualsiasi informazione pertinente relativa alla procedura e ai principi di utilizzo, con mezzi automatizzati, di dati personali al fine di ottenerne un determinato risultato", a cui si devono aggiungere le informazioni relative all'importanza e alle conseguenze previste del trattamento controverso per l'interessato. Il diritto di ottenere "informazioni significative sulla logica utilizzata" in un processo decisionale automatizzato deve essere inteso come un diritto alla spiegazione della procedura e dei principi concretamente applicati per utilizzare, con mezzi automatizzati, i dati personali dell'interessato al fine di ottenerne un risultato specifico. "Non può soddisfare tali requisiti né la semplice comunicazione di una formula matematica complessa, come un algoritmo, né la descrizione dettagliata di tutte le fasi di un processo decisionale automatizzato, in quanto nessuno di questi metodi costituirebbe una spiegazione sufficientemente concisa e comprensibile" (par. 59). La ponderazione tra diritto di accesso e tutela del segreto commerciale deve essere effettuata caso per caso dall'autorità di controllo o dal giudice competente. Uno Stato membro, quindi, non può prevede una disposizione che escluda, di regola, il diritto di accesso dell'interessato, previsto all'articolo 15 del GDPR, qualora tale accesso comprometta un segreto commerciale o un segreto aziendale del titolare del trattamento o di un terzo, stabilendo cioè in modo definitivo, a priori, l'esito della ponderazione dei diritti e degli interessi in gioco.

12 Rispetto all'obbligo informativo nei confronti anche dei rappresentanti dei lavoratori, da adempiere prima della messa in servizio o utilizzo del sistema di IA nell'ambiente di lavoro, cfr. art. 26, par. 7.

13 Per una definizione di sistemi di monitoraggio automatizzati, cfr. la Direttiva (UE) 2024/2831 sul lavoro mediante piattaforme digitali, in cui tali sistemi sono qualificati come "sistemi utilizzati per effettuare, o che sostengono, il monitoraggio, la supervisione o la valutazione, tramite strumenti elettronici, dell'esecuzione del lavoro delle persone che svolgono un lavoro mediante piattaforme digitali o delle attività svolte nell'ambiente di lavoro, anche raccogliendo dati personali" (art. 1, par. 1, lett. h).

stabilisce un obbligo di informazione del lavoratore dell'utilizzo dell'IA "nei casi e con le modalità di cui all'art. 1-bis" (art. 11, comma 2).

Gli orientamenti della Commissione confermano altresì che un sistema di IA, per essere definito tale ai fini del Regolamento, non debba necessariamente presentare adattabilità dopo la diffusione, intesa come capacità di autoapprendimento dopo l'installazione e durante l'uso, con possibile modifica dei pesi con cui le diverse variabili influenzano l'output e conseguente comportamento auto-evolutivo della macchina. Tale caratteristica può, invece, rilevare ai fini della categorizzazione del rischio: in forza dell'art. 6, par. 1, dell'AI Act, infatti, è considerato ad alto rischio un sistema destinato ad essere utilizzato come componente di sicurezza di un prodotto o che è esso stesso prodotto e che ai sensi della normativa di armonizzazione UE elencata nell'allegato I, tra cui rientra il nuovo Regolamento Macchine, debba essere sottoposto a una valutazione di conformità da parte di terzi. Tale è, in particolare, stando al citato Regolamento Macchine, un sistema con comportamento integralmente o parzialmente auto-evolutivo, quindi dotato di adattabilità durante l'uso, volto a garantire una funzione di sicurezza in un macchinario. Si pensi a un sistema di IA che rileva la presenza di operatori od ostacoli vicino a una pressa o a un veicolo industriale/robot che blocca/adatta il movimento o i parametri di funzionamento in caso di pericolo o per evitare collisioni; a un sistema di IA che adatta il funzionamento di un escavatore al terreno per ridurre o eliminare alcuni rischi (ad esempio di vibrazione); a un sistema di manutenzione predittiva che identifica guasti imminenti; a un sistema che modifica forza, velocità, ritmo di movimento di un braccio robotico in un impianto di produzione in base ai dati che raccoglie in tempo reale. Solo in presenza di adattabilità durante l'uso, tali sistemi, pur rientrando nella definizione di sistemi di IA, saranno sottoposti al principale quadro di garanzie previste dall'AI Act, dedicato ai sistemi ad alto rischio.

Ancora, gli orientamenti della Commissione ribadiscono come ai fini della definizione di sistema IA non rilevi il tipo di approccio che consente l'inferenza, quanto la capacità di inferenza stessa. La tecnica che permette di dedurre l'output può essere, pertanto, sia di ragionamento automatico (*logic-based*), sia di apprendimento automatico. Lo stesso *machine learning* si può articolare in varie tipologie, dal supervisionato al non supervisionato, all'auto-supervisionato, come sottocategoria del secondo, fino all'auto-apprendimento per rinforzo.

Ciò che conta, per la Commissione, è che il sistema sia in grado, nel processo inferenziale, di gestire relazioni e modelli complessi nei dati.

A rimanere esclusi dalla definizione sono i sistemi "che non trascendono l''elaborazione dati di base'", come quelli che consentono di migliorare la performance computazionale, ma non di adeguare il modello decisionale; le applicazioni software standard per fogli di calcolo (privi di applicazioni di IA) o i sistemi di gestione di database che permettono di ordinare o filtrare i dati in base a criteri specifici; ancora, i sistemi di previsione semplici, di analisi descrittiva o di stima statica.

L'elenco esemplificativo degli esclusi non consente, d'altra parte, di segnare il confine della fattispecie. Del resto, il confine non può che rimanere poroso, se a guidarne la demarcazione è il parametro della complessità. È la stessa Commissione a segnalare come l'ampia ed elastica definizione in esame non permetta di determinare automaticamente o di elencare in modo esaustivo i sistemi che vi rientrano: la qualificazione richiede una verifica caso per caso[14].

Certo: se il criterio della complessità non ci consente di predeterminare in modo automatico l'applicabilità della definizione, è proprio la complessità del fenomeno da regolare che spiega la necessità dell'approccio basato sul rischio, come tecnica normativa volta a garantire l'effettività dei diritti fondamentali (Loi 2001).

Sul punto, tuttavia, è necessaria una precisazione.

14 Le problematiche derivanti da una definizione molto ampia e dai confini porosi sono state segnalate con forza dai partecipanti alla consultazione pubblica sulle linee guida. Nel report di sintesi pubblicato dalla Commissione, si sottolinea, in particolare che "Le risposte indicano un ampio consenso sulla necessità di approfondire ulteriormente le definizioni di 'adattività', 'inferenza' e 'autonomia'. Molti portatori di interesse hanno espresso preoccupazione per il fatto che le definizioni attuali potrebbero includere, inavvertitamente, anche sistemi software tradizionali, come quelli basati su regole o modelli statistici di base, che non presentano le caratteristiche proprie dell'Intelligenza artificiale", Commissione europea (2025, 21).

Come evidenzia la Commissione nei propri orientamenti sulla definizione di sistema di IA, l'approccio *risk-based* trova implementazione, all'interno del Regolamento sull'IA, a partire da una distinzione per categorie di rischio (inaccettabile, alto e non-alto), a cui corrispondono previsioni di divieto ovvero l'obbligo di rispettare il principale *corpus* di prescrizioni stabilite dal Capo III[15]. La tecnica di giuridificazione del rischio utilizzata è, d'altra parte, costituita dall'individuazione di insiemi di pratiche d'uso e di ipotesi di deroga tipizzate, costruite a livello normativo con enunciati troppo generici o di dubbia interpretazione, tali da rendere ancora una volta le fattispecie di non chiara demarcazione (Mantelero e Peruzzi 2024). Gli orientamenti sulle pratiche vietate hanno certamente sciolto alcuni interrogativi, come quello sui sistemi di inferenza delle emozioni, sorto a fronte di un contrasto di lettura tra quanto indicato nel Cons. n. 18 e quanto riportato sulla pagina istituzionale della Commissione (ora è chiaro che i sistemi che rilevano il dolore o la stanchezza non inferiscono emozioni). I 434 paragrafi in cui si articolano le linee guida dedicate (solo) all'art. 5 sono, d'altra parte, indicativi della problematicità della lettura e del tracciamento degli ambiti di applicazione delle regole. A questi orientamenti si aggiungeranno quelli – attesi, a norma dell'art. 6, par. 5, entro il 2 febbraio 2026 – relativi all'applicazione delle deroghe alla classificazione ad alto rischio dei sistemi elencati nell'allegato III. Tali indicazioni saranno tanto più rilevanti se si considera che il fornitore, in molti casi, come quello dei sistemi utilizzati per la selezione e la gestione del personale, effettua la classificazione sulla base di un processo interno di self-assessment, dovendo sì documentare la valutazione, ma non anche depositare tale documentazione al momento della registrazione del sistema nella banca dati UE, posto che a essere inserita, tra le informazioni da allegare, è solo l'indicazione delle condizioni che, in base all'art. 6, par. 3, hanno consentito di non ritenere detto sistema ad alto rischio. La documentazione è conservata internamente dal fornitore e messa a disposizione delle autorità competenti, su richiesta. L'autorità di vigilanza del mercato (per l'Italia, l'Autorità per la cybersicurezza nazionale, secondo quanto previsto dal D.D.L. governativo[16]), qualora abbia motivi sufficienti per ritenere che la classificazione non sia corretta, effettua una valutazione del sistema ed eventualmente richiede al fornitore di adottare le misure necessarie per garantire la sua conformità (art. 80).

Come già segnalato, a norma dell'art. 6, par. 3, il rischio significativo di danno – cui segue la classificazione ad alto rischio – è ritenuto escluso laddove il sistema non influenzi materialmente il risultato del processo decisionale, ovvero sempre presente qualora il sistema effettui profilazione di persone fisiche.

La tipizzazione delle condizioni che escludono il rischio, effettuata dalla disposizione, solleva non pochi dubbi interpretativi, soprattutto in rapporto ad alcune aree critiche d'uso elencate dall'allegato III. Il rischio significativo di danno sarebbe, ad esempio, da escludere, qualora il sistema fosse solo volto a rilevare lo scostamento di una valutazione umana

15 Per i sistemi non classificabili come ad alto rischio trovano applicazione, se del caso, le garanzie di trasparenza di cui al Capo IV ovvero può essere prevista l'applicazione volontaria delle garanzie del Capo III tramite codici di condotta (elaborabili anche da parte di organizzazioni di *deployer*, quindi anche di rappresentanza datoriale, insieme con rappresentanti dei portatori di interessi, quindi anche sindacati, a norma dell'art. 95). Come chiarisce il Cons. n. 166, "è importante che i sistemi di IA collegati a prodotti che non sono ad alto rischio in conformità del presente regolamento e che pertanto non sono tenuti a rispettare i requisiti stabiliti per i sistemi di IA ad alto rischio siano comunque sicuri al momento dell'immissione sul mercato o della messa in servizio. Per contribuire a tale obiettivo, sarebbe opportuno applicare come rete di sicurezza il Regolamento (UE) 2023/988" relativo alla sicurezza generale dei prodotti.

16 La scelta di affidare il ruolo a una agenzia soggetta alle funzioni di indirizzo e coordinamento della Presidenza del Consiglio, ovvero, come avvenuto in Spagna, a una nuova agenzia di vigilanza ad hoc, comunque presieduta dal titolare della Segreteria di stato per la digitalizzazione, è consentita dall'AI Act che non richiede il carattere indipendente dell'*authority*, ma solo che eserciti i poteri "in modo indipendente, imparziale e senza pregiudizi, in modo da salvaguardare i principi di obiettività delle loro attività e dei loro compiti" (art. 70). Sul punto, cfr. le considerazioni di Falletta e Marsano (2024). Come spiegano Schmidl e Rohner (2025, 1025), la differenza tra la "completa indipendenza" richiesta dal GDPR alle relative *authority* (e che anche il Pe aveva invero inserito pur solo nel considerando n. 77 nel suo testo di compromesso dell'AI Act) e l'indipendenza prevista dal Regolamento sull'IA è sostanziale: "mentre la completa indipendenza delle autorità per la protezione dei dati richiede che esse possano 'scegliere e disporre di personale proprio', soggetto solo alla 'direzione esclusiva del membro o dei membri' dell'autorità per la protezione dei dati (cfr. articolo 52, paragrafo 5, del GDPR), e che debbano disporre di un 'bilancio annuale separato e pubblico' (articolo 52, paragrafo 6, del GDPR), requisiti analoghi non si trovano nella legge sull'AI".

già completata rispetto a uno schema prestabilito, potendo procedere a una sua sostituzione o determinando un condizionamento della stessa solo con adeguata revisione umana. L'indicazione interroga sui termini con cui la presenza della revisione umana in un processo decisionale automatizzato sia in grado di escludere l'alto rischio (quando, cioè, rilevi e si possa ritenere di misura adeguata).

Parimenti il rischio è escluso se il sistema di IA esegue un compito procedurale limitato (ad esempio classificazione di documenti per categorie), migliora il risultato di un'attività umana precedentemente completata (ad esempio allineamento del testo di un documento già redatto a un determinato stile), ovvero esegue un compito solo preparatorio per una valutazione pertinente ai casi d'uso di cui all'allegato III, quindi ad esempio ai fini della selezione o gestione del personale. Il Cons. n. 53 segnala, tra i possibili compiti preparatori, "soluzioni intelligenti per la gestione dei fascicoli, che comprendono varie funzioni quali (...) il collegamento dei dati ad altre fonti di dati". La delimitazione della fattispecie andrebbe però meglio chiarita, perché il generico enunciato normativo non consente di capire se o quando, ad esempio, uno *screening* di candidature effettuato da un sistema di IA, indicato tra le finalità d'uso critiche dell'allegato III (sistemi utilizzati "per filtrare le candidature e valutare i candidati", ma anche sistemi "destinati a essere utilizzati per adottare decisioni riguardanti le condizioni dei rapporti di lavoro, la promozione"), possa integrare la condizione di deroga alla classificazione ad alto rischio, in quanto compito prodromico a restringere la rosa di candidati da sottoporre a una valutazione umana. Il tema è tanto più rilevante considerati i rischi di discriminazione che tali processi di preselezione possono implicare. Si pensi a un sistema addestrato con tecniche di apprendimento supervisionato sulla base di un set di dati di training non sufficientemente rappresentativi: imparerà a penalizzare i candidati (o le candidate…) che non presentano i pattern ricorrenti nel gruppo che risulta sovrarappresentato e, quindi, agli 'occhi' della macchina, statisticamente più adatto all'obiettivo (c.d. *proxy discrimination*); parimenti, una discriminazione (indiretta) potrebbe annidarsi nell'applicazione di tecniche di apprendimento supervisionato (per addestrare un modello funzionale, ad esempio, al ranking di candidati a una promozione o a un bonus) in cui la selezione delle variabili indipendenti collegate alla variabile target (la c.d. etichetta) includa fattori apparentemente neutri, ma ad impatto differenziato, come la tipologia del contratto (full-time/part-time) o il numero di giorni di assenza (Gosetti *et al.* 2024). Ad ogni modo, quantomeno nell'esempio summenzionato, pare difficile escludere che il processo, riducendo l'ambito della scelta, non influenzi materialmente la decisione umana finale. Soprattutto, a risolvere a monte molte questioni interpretative poste dalla costruzione delle ipotesi di deroga dovrebbe essere l'applicabilità dell'ipotesi di conferma del rischio sempre stabilita dal par. 3, ossia la profilazione di persone fisiche. Per la definizione, l'AI Act rinvia a quanto stabilito nel GDPR, che qualifica la fattispecie come un trattamento automatizzato finalizzato a valutare, ai fini di analisi e/o previsioni, aspetti e caratteristiche personali individuali, come il rendimento, l'ubicazione, l'affidabilità, il comportamento. Essendo molto probabile che tale condizione ricorra in caso di utilizzo di processi automatizzati di monitoraggio o decisionali sul lavoro, può ritenersi tendenzialmente da escludere una classificazione come 'a basso rischio' dei sistemi destinati a essere utilizzati nell'esercizio di prerogative datoriali nei confronti dei lavoratori.

Necessità e declinazione dell'approccio risk-based. Gli obblighi di valutazione e gestione del rischio del datore di lavoro nell'AI Act, come deployer e fornitore

Come messo in luce dalla dottrina, la tecnica normativa dell'approccio basato sul rischio, nel tradurre "una prospettiva procedurale del diritto", rappresenta una "modalità di reazione del diritto ad un elevato grado di differenziazione" e di complessità del reale, tale da impedire una predeterminazione rigida e standard dei contenuti delle misure di tutela (Loi 2001, 47-48). Si tratta, pertanto, di un modello regolatorio di risposta all'inadeguatezza che può presentare la regolazione diretta per misure di dettaglio 'nominate' di fronte alla dinamicità evolutiva dei fattori di pericolo e alla modularità degli elementi che li caratterizzano.

Se, come insegna il diritto della sicurezza sul lavoro, la tecnologia è sempre stata una determinante chiave di questa complessità, nel caso dell'IA, soprattutto del *machine learning*, la problematica si intensifica e acuisce, perché le peculiarità adattive e auto-evolutive della

fonte di potenziale pericolo, nonché la variabilità della sua possibile autonomia d'azione fanno dipendere l'effettività delle tutele da un necessario processo iterativo di valutazione e gestione dei rischi. Sempre le caratteristiche dell'IA – ossia la dipendenza dai dati per l'applicazione delle tecniche di inferenza, l'opacità del processo di ottenimento dell'output e la capacità di quest'ultimo di influenzare gli ambienti fisici e virtuali – impongono che il modello di controllo del rischio si muova sul duplice piano della governance dei dati e della verifica periodica di impatto dei processi automatizzati. Ciò è tanto più vero all'aumentare della profondità delle reti neurali e dell'apprendimento automatico (profondità che sta alla base, ad esempio, dello sviluppo dei modelli di GPAI[17]), considerato che in tal caso l'investigazione può realizzarsi solo attraverso un'osservazione dall'esterno del modello addestrato, nel suo funzionamento e nei risultati che consegna (Italiano *et al.* 2022).

La scelta di metodo, nella costruzione del sistema integrato delle fonti, si confronta, d'altra parte, nella prospettiva lavoristica con la necessità di "delimitazione di zone di indisponibilità, costituite dai diritti sociali fondamentali" (Loi 2001, 48). In altre parole, se il paradigma procedurale dell'approccio *risk-based* si presenta funzionale a garantire l'effettività dei diritti nel descritto scenario di complessità, la definizione di tali diritti deve essere presidiata a livello di fonti legislative e costituzionali, in modo che solo queste "rappresentino l'adeguato spazio giuridico in cui effettuare il bilanciamento degli interessi coinvolti" (Loi 2001, 68-69). È in tale spazio che deve essere, cioè, individuato il grado di tutela dei diritti, venendo affidata all'attore privato, titolare dell'obbligo di valutazione e gestione, solo la scelta delle misure tecniche e organizzative, ossia dei metodi e degli strumenti, che meglio lo garantiscono nello specifico contesto d'incidenza della fonte di pericolo.

Il modello previsto dalla Direttiva 89/391/CEE, concernente la salute e sicurezza durante il lavoro, rappresenta l'esempio di matrice lavoristica della convivenza di questi due paradigmi (Loi 2001). È un esempio che salda il grado di tutela dei diritti a un principio di massima sicurezza tecnologica[18], sganciandolo "da considerazioni di carattere puramente economico" (Cons. n. 13), e rispetta la necessaria proiezione collettiva delle tutele richiesta dalla materia, prevedendo il coinvolgimento di rappresentanze specializzate nel processo di valutazione e gestione del rischio (v. *infra*). Significativamente, all'interno della direttiva 89/391 la combinazione dei due paradigmi incide sulla stessa declinazione dell'approccio basato sul rischio, essendo questo informato dalla prospettiva della c.d. prevenzione primaria, ossia di un modello di tutela dei diritti fondamentali che incorpora "un bilanciamento già intervenuto per opera del Legislatore", prevedendo "la tensione alla radicale eliminazione del rischio" e "la copertura surrogatoria in senso limitativo o riduttivo della dose di rischio ineliminabile" (Buoso 2020, 48-49). In tal senso, il paradigma procedurale è segnato da una precisa sequenza di obiettivi: prima l'eliminazione del rischio; in subordine, la riduzione al minimo dei rischi ineliminabili (Lazzari e Pascucci 2024).

Si è già potuto in altra sede riflettere su quanto l'importanza della grammatica dei diritti fondamentali trovi conferma anche nel contesto dei due principali plessi regolativi coinvolti nell'integrazione tra fonti (AI Act e GDPR) e quanto la declinazione del paradigma procedurale *risk-based* all'interno di tali sistemi possa trovare allineamento con lo standard tracciato dal quadro prevenzionistico (Peruzzi 2024; cfr. al riguardo anche le diverse riflessioni di Gaudio 2024 e Treu 2025).

Pare, d'altra parte, utile tornare sull'incidenza dell'approccio basato sul rischio sulla figura del *deployer*-datore nel contesto normativo dell'AI Act. In forza dell'art. 26, tale soggetto è tenuto a dare applicazione alle misure di sorveglianza umana indicate dal fornitore, finalizzate espressamente, ai

17 Cfr. Orofino (2025). Come evidenzia la dottrina, "il successo delle reti neurali profonde nel riconoscimento delle immagini ha condotto a estenderne l'uso anche ai compiti linguistici" ed "è proprio nell'ambito linguistico che le reti neurali profonde hanno consentito di ottenere i risultati più significativi". Si fa riferimento a "i grandi modelli linguistici (*Large Language Model* o LLM)" ossia "enormi reti neurali pre-addestrate, la cui realizzazione è stata resa possibile dalla combinazione di tre fattori: enormi raccolte di testi da utilizzare per l'addestramento; altissima potenza di calcolo; nuove tecnologie per la realizzazione delle reti". "In generale, il funzionamento degli LLM resta ancora assai misterioso nel senso che non esiste una precisa teoria che spieghi perché tali sistemi forniscano le sorprendenti prestazioni di cui sono capaci" (Sartor *et al.* 2024, 252-256).

18 Rispetto alla possibilità di collocare, nell'ambito di operatività di tale principio, l'introduzione della tecnologia avanzata, nella forma di Intelligenza artificiale e robotica, cfr. di recente Faioli (2025).

sensi dell'art. 14, in sequenza, a prevenire o ridurre al minimo i rischi per la salute e sicurezza o i diritti fondamentali; ed è altresì chiamato a monitorare debitamente il funzionamento del sistema (anche attraverso la registrazione e conservazione dei log di tracciamento), dovendo sospenderne l'uso, qualora abbia ragione di ritenere, a fronte del monitoraggio, che nonostante il rispetto delle istruzioni ricevute il sistema presenti un rischio per la salute e sicurezza e i diritti fondamentali. Se un obbligo di valutazione di impatto sui diritti fondamentali (c.d. FRIA) è espressamente prescritto, dall'art. 27, solo per gli organismi di diritto pubblico o gli enti privati che forniscono servizi pubblici (come ospedali, scuole, banche e compagnie di assicurazione), la riflessione sull'importanza del paradigma *risk-based* in un contesto di utilizzo dell'IA invita, infine, a considerare una sua introduzione (volontaria) trasversale nelle posizioni d'obbligo datoriali ai fini di un loro corretto adempimento (Mantelero 2024; Peruzzi 2024)[19].

Al di là di questi profili, si deve d'altra parte tener conto della possibilità che in capo al datore di lavoro gravi la posizione non solo di *deployer*, ma anche di fornitore, con tutti i relativi obblighi in termini di valutazione e gestione dei rischi del sistema di IA in questione, se classificabile come ad alto rischio.

Anzitutto, a norma dell'art. 3, n. 3, è fornitore non solo chi sviluppa un sistema di IA o un modello di IA per finalità generali (GPAI), ma anche chi li fa sviluppare, per poi commercializzarli o anche solo metterli in servizio (ossia consegnarli al *deployer* per uso proprio), purché apponga il proprio nome (ad esempio OpenAI, L.P.) o marchio (ad esempio ChatGPT) (Bassini 2024; il marchio può essere anche non registrato, cfr. Feiler e König 2025). In tal senso, il datore può figurare al contempo come fornitore e *deployer* del sistema, laddove utilizzi software sviluppati internamente o fatti sviluppare e in ipotesi anche commercializzati, a cui abbia apposto il proprio nome o marchio. Il cumulo di posizioni si configura anche se il datore sviluppi o faccia sviluppare un sistema che integra un modello di IA (c.d. fornitore a valle), indipendentemente dal fatto che il modello sia fornito dallo stesso o da terzi[20]. Il datore potrebbe quindi figurare, al contempo, fornitore a valle e *deployer* del sistema, nel momento in cui dapprima sviluppi o faccia sviluppare (con suo nome o marchio) un sistema che integra un modello di GPAI, quale ad esempio un modello di LLM per l'analisi del linguaggio naturale (ad esempio GPT-4, Bert), adattandolo o 'accordandolo' (c.d. *fine-tuning*) per eseguire come compito l'analisi del contenuto dei curricula ricevuti con selezione dei migliori candidati, quindi lo applichi direttamente per tale finalità d'uso[21]. Il *fine-tuning* può avvenire, ad esempio, con ri-addestramento supervisionato del modello, con nuovi dati specifici per il compito e l'ambito/settore interessato (ad esempio con i dati storici dei CV ricevuti dall'azienda in precedenza). Resta ferma, anche in questo, ai fini della qualificazione del soggetto come fornitore, la messa in servizio (se non anche commercializzazione) di un sistema di IA autonomo (quindi con una propria struttura applicativa, ad esempio una interfaccia utente, integrazione con basi di dati, flussi decisionali automatizzati) sotto il proprio nome o marchio.

È importante considerare, tuttavia, che il *deployer* è soggetto agli obblighi del fornitore, a norma dell'art. 25, anche laddove apporti a un sistema di IA ad alto rischio – già immesso sul

19 La prospettiva di una applicazione trasversale dell'approccio *risk-based* trova conferma nelle linee guida elaborate a livello interno dall'Agid per la Pubblica amministrazione e dal Ministero del Lavoro per il mondo del lavoro. In particolare, nella *Bozza di linee guida per l'adozione di IA nella pubblica amministrazione*, adottata e sottoposta a consultazione pubblica dall'AGID con la determinazione n. 17 del 17 febbraio 2025, si propone, accanto alla FRIA e DPIA, anche l'AIIA (*Artificial Intelligence Impact Assessment*). Nelle *Linee guida per l'implementazione dell'IA nel mondo del lavoro*, presentate dal Ministero del Lavoro il 14 aprile 2025 con apertura di una consultazione pubblica, si evidenzia la necessità di una valutazione di impatto – con verifica dei rischi e implementazione delle misure di mitigazione – sull'occupazione e transizione lavorativa, privacy e protezione dei dati, sicurezza e condizioni di lavoro, discriminazione e *bias* algoritmici (p. 40).

20 Laddove il modello sia sviluppato dallo stesso fornitore (è il caso di OpenAI e ChatGpt), il soggetto sarà gravato anche di tutti gli obblighi di cui al Capo V, se si tratta di modello di GPAI. Cfr. il codice di buone pratiche sull'IA per finalità generali e gli Orientamenti in materia (C(2025) 5045 final) pubblicati dalla Commissione a luglio 2025.

21 Cfr. il sistema proposto da Sareddy e Farhan (2025), basato sull'integrazione di un modello di LLM (BERT), modificato nella sua architettura (Nish-BERT) e quindi adattato per lo screening dei CV attraverso tecniche di clusterizzazione e di apprendimento supervisionato.

mercato o fornitogli – una modifica sostanziale[22] che non incida su tale classificazione; ovvero anche laddove modifichi la finalità prevista di un sistema di IA, sia anche un sistema per finalità generali (GPAI)[23], in modo tale da richiederne una classificazione ad alto rischio. Il tema è di estrema rilevanza, posto che in questo caso l'assunzione della qualifica di fornitore non avviene a fronte dell'apposizione di nome e marchio al prodotto (per quanto poi richiesta, in forza della conseguente applicazione dell'art. 16, par. 1, lett. b)[24]. Se ancora una volta la fattispecie della 'modifica sostanziale' dovrà essere chiarita dagli orientamenti della Commissione ai sensi dell'art. 96, pare utile richiamare sul punto l'art. 43, par. 4, secondo cui non costituiscono modifiche sostanziali quelle apportate al sistema e alle sue prestazioni dal processo di apprendimento continuo sviluppato dalla macchina dopo l'immissione sul mercato o la messa in servizio, purché predeterminate dal fornitore al momento della valutazione iniziale di conformità (v. Cons. 128).

Il profilo da ultimo esaminato richiede di porre particolare attenzione all'ipotesi in cui il *deployer* di un sistema di GPAI venga gravato anche degli obblighi propri del fornitore per averne modificato la finalità prevista.

La riflessione impone a monte due considerazioni. Anzitutto, a poter essere messo in servizio da un fornitore e utilizzato poi da un *deployer* è solo un sistema di GPAI (ad esempio ChatGPT) e non anche un modello di GPAI (ad esempio GPT-5). In altre parole, "un modello GPAI non può essere messo in servizio in modo autonomo. Solo una volta che il modello GPAI è stato integrato in un sistema di Intelligenza artificiale, esso può essere messo in servizio" (Voigt e Hullen 2024, 26); parimenti, l'AI Act non si applica o riguarda gli utilizzatori di modelli GPAI, ma solo i *deployer* di sistemi di IA: del resto 'il 'funzionamento' di un modello di IA avviene sempre tramite un sistema di IA – ed è per quest'ultimo che possono essere previsti obblighi specifici a carico degli utilizzatori, distributori o importatori (in particolare nel caso di sistemi di IA ad alto rischio)" (Voigt e Hullen 2024, 148). La seconda considerazione attiene, invece, ai termini in cui si può individuare una modifica di finalità prevista in caso di sistema di GPAI, tale da integrare una modifica sostanziale del sistema. Al riguardo, la dottrina ha messo in luce che "non tutte le finalità potenziali di un sistema GPAI devono essere intese come finalità previste, ma solo quelle che il fornitore ha già specificato come campi di applicazione"; "la nozione di 'finalità generale' si riferisce alle capacità oggettive, indipendentemente dalle intenzioni del fornitore. La finalità prevista, invece, si riferisce esclusivamente alle intenzioni soggettive del fornitore. Questa distinzione è supportata anche dal fatto che l'uso improprio ragionevolmente prevedibile non rientra nella finalità prevista. Di conseguenza, la finalità prevista non dovrebbe dipendere dalla conoscenza del fornitore di come il sistema di IA potrebbe essere utilizzato, ma solo dalle finalità che il fornitore ha volontariamente assegnato al sistema" (Voigt e Hullen 2024, 26).

Si pensi, pertanto, all'ipotesi di un datore di lavoro che, come *deployer*, utilizzi ChatGPT e che, accedendo alla funzione che consente di 'customizzare' una versione della applicazione per una specifica finalità, decida di impostarne vincoli di contenuto tramite istruzioni strutturate (c.d. *prompt engineering*, a sua volta realizzabile con

22 La modifica è definita sostanziale, a norma dell'art. 3, n. 23, quando avviene a seguito della immissione sul mercato o messa in servizio del sistema, non è prevista o programmata nella valutazione iniziale di conformità effettuata dal fornitore e ha l'effetto di incidere sulla conformità del sistema di IA ai requisiti di cui al capo III, sezione 2, ovvero comporta una modifica della finalità prevista per la quale il sistema di IA è stato valutato. Come evidenzia Bassini (2024, 133), "il mutamento di qualificazione, che comporta l'assunzione dei relativi obblighi, non è però unilaterale, ma risulta 'biunivoco': in questi casi, infatti, l'originario fornitore che ha immesso sul mercato o messo in servizio il sistema di Intelligenza artificiale non sarà più considerato fornitore di quel determinato sistema, ferma restando l'esigenza di cooperazione con il 'nuovo' fornitore e la necessità di fornire tutte le informazioni necessarie".

23 A norma dell'art. 3, si tratta di "un sistema di IA basato su un modello di IA per finalità generali e che ha la capacità di perseguire varie finalità, sia per uso diretto che per integrazione in altri sistemi di IA".

24 Come evidenziano Riede e Talhoff (2025, 529), "se un tale cambio di ruolo avviene involontariamente, si crea un rischio significativo di responsabilità per il 'nuovo fornitore', poiché è altamente improbabile che adempia inconsapevolmente ai propri obblighi in quanto fornitore di un modello di IA ad alto rischio. Tutti gli operatori nella catena del valore dell'IA che apportano qualsiasi tipo di modifica ai sistemi di IA dovrebbero monitorare attentamente la conformità all'art. 25 per evitare di incorrere in responsabilità come '*deemed provider*'".

il supporto dell'IA[25]), con limitazione del dominio d'uso e integrazione della base di conoscenza con allegazione di fonti informative dedicate, ovvero anche collegamenti a banche dati aziendali tramite API. Ci si può chiedere, alla luce di quanto sopra esposto, se la personalizzazione, volta ad esempio a richiedere all'applicazione di valutare dei CV e predisporre un ranking dei candidati, implichi una modifica della finalità d'uso prevista del sistema – ad esempio da assistente generale per interazione conversazionale a selezionatore del personale – e, quindi, una modifica sostanziale tale da determinare l'applicazione al *deployer* anche degli obblighi stabiliti per il fornitore dall'art. 16[26]. Il passaggio richiede, come detto, che la modifica determini o confermi una classificazione del sistema come ad alto rischio a norma dell'art. 6, una classificazione che richiede, d'altra parte, una verifica non solo dell'inclusione della pratica d'uso tra quelle critiche elencate dall'allegato III (ad esempio, analizzare o filtrare le candidature e valutare i candidati), ma anche della sussistenza di un rischio effettivo per i diritti fondamentali, esclusa, come detto, laddove l'output non incida sostanzialmente sull'esito del processo decisionale, ma sempre presente laddove il sistema effettui una profilazione di persone fisiche (come avviene nel caso in esame).

IA e contrattazione collettiva: le esperienze in Italia

Si è più volte evidenziata la necessità di dedicare specifica attenzione al ruolo delle Parti sociali, nella costruzione della dimensione collettiva delle tutele che lo stesso AI Act richiama laddove prevede che lo standard minimo di garanzie predisposto possa essere incrementato a favore dei lavoratori tramite contratto collettivo (v. *supra*). Sempre nel contesto del Regolamento, a parte il (mero) obbligo informativo alle rappresentanze imposto al datore-*deployer* prima della messa in servizio/utilizzo di un sistema ad alto rischio (art. 26, par. 7), si prevede il possibile ruolo delle Parti sociali nell'elaborazione dei codici di condotta e dei relativi sistemi di governance (Ciucciovino 2024), intesi a declinare i processi di alfabetizzazione sull'IA del personale[27], nonché l'applicazione volontaria delle garanzie ai sistemi a basso rischio ovvero di requisiti supplementari, a promozione di processi inclusivi e partecipati, ai sistemi ad alto rischio (art. 95; Cons. n. 165). Varchi d'accesso per il coinvolgimento delle rappresentanze si aprono nel raccordo con le fonti lavoristiche (si pensi, al ruolo del Rappresentante dei lavoratori per la sicurezza (RLS) nella valutazione e gestione dei rischi su salute e sicurezza ovvero delle rappresentanze ai fini dell'installazione di strumenti che consentono il controllo tecnologico a distanza) e non (il riferimento è anzitutto alla consultazione delle rappresentanze che può essere richiesta, se del caso, nella valutazione di impatto, c.d. DPIA, dalla normativa sulla protezione dei dati personali). Una specifica declinazione collettiva delle tutele rispetto alla gestione algoritmica è, infine, predisposta dalla Direttiva Piattaforme 2024/2831 (Aimo 2024), per il proprio ambito di applicazione, non a caso guardata come modello

25 Come spiega la dottrina "Negli LLM più potenti (come GPT-4) è possibile fornire istruzioni di input (prompt) di lunghezza considerevole (circa 50 pagine di testo). Pertanto, tali istruzioni possono comprendere descrizioni dettagliate del risultato da ottenere e del procedimento da seguire, esempi di casi in cui il compito sia stato eseguito correttamente ecc." (Sartor *et al.* 2024, 255).

26 Come segnalano Riede e Talhoff (2025, 533), "per questi motivi, gli utilizzatori di sistemi AI saranno generalmente invitati a non utilizzarli in contesti ad alto rischio, a meno che il sistema AI ad alto rischio in questione non sia stato specificamente progettato per tale finalità dal fornitore".

27 Il profilo è tutt'altro che secondario, soprattutto se si legge il Cons. n. 20, ove si afferma che il processo include le conoscenze necessarie per comprendere in che modo le decisioni adottate con l'assistenza dell'IA possano incidere sulle persone interessate. L'obbligo di alfabetizzazione, previsto dall'art. 4 AI Act, entrato in applicazione il 2 febbraio 2025, riguarda tutti i *deployer* di sistemi di IA, siano essi ad alto o basso rischio. Come chiarito dalla Commissione europea sul proprio sito, nella sezione *AI Literacy - Questions & Answers*, l'alfabetizzazione, a norma dell'art. 3 AI Act, deve includere "le competenze, le conoscenze e la comprensione che consentono (...) di procedere a una diffusione informata dei sistemi di IA, nonché di acquisire consapevolezza in merito alle opportunità e ai rischi dell'IA e ai possibili danni che essa può causare". Il contenuto dell'obbligo è d'altra parte modulare e dipende dalle caratteristiche dell'organizzazione, dello strumento introdotto e dei relativi rischi, dal tipo di lavoratori coinvolti e dalle loro competenze: "in molti casi" – segnala la Commissione – "affidarsi semplicemente alle istruzioni per l'uso dei sistemi di Intelligenza artificiale o chiedere al personale di leggerle potrebbe risultare inefficace e insufficiente", "non esiste un approccio valido per tutti (...). I requisiti che deve presentare la formazione dipendono dal contesto concreto". Il tema dell'alfabetizzazione del personale è oggetto di particolare attenzione sia nelle *Linee guida per l'utilizzo dell'AI nella Pubblica amministrazione*, sia in quelle per il mondo del lavoro presentate dal Ministero del Lavoro (v. *supra*).

per un possibile rafforzamento delle garanzie per i lavoratori in materia, a livello UE (v. *supra*).

Rinviando ad altra sede una riflessione più complessiva sul punto, pare utile qui richiamare alcune esperienze nazionali, indicative del 'fermento' che si sta registrando a livello di contrattazione collettiva in Italia.

In principio, si può ricordare il noto Accordo Tim del 29 aprile 2020 – sottoscritto ai sensi dell'art. 4 St. Lav. con SLC-CGIL, FISTel-CISL, UILCom-UIL, UGL Telecomunicazioni, unitamente al Coordinamento nazionale RSU – volto a disciplinare l'uso della tecnologia 'Routing Evoluto' Afiniti, in grado di creare abbinamenti tra i clienti chiamanti e gli operatori telefonici, utilizzando sistemi di IA (Imberti 2023). Ancora, nel contesto dell'Accordo territoriale sottoscritto il 14 giugno 2022 dalle sigle di Confcooperative e Cgil, Cisl e Uil competenti per la provincia di Cuneo – finalizzato a disciplinare l'allineamento al CCNL per i lavoratori dipendenti da aziende cooperative di trasformazione di prodotti agricoli e zootecnici e lavorazione prodotti alimentari – sono condivisi alcuni 'assunti fondamentali' a guida dell'introduzione di robots/cobots e/o IA nelle singole realtà aziendali. Nello specifico, si evidenzia la necessità, sul piano dei contenuti, di garantire il controllo umano, "con riduzione della possibilità di adozione di misure o di decisioni in modo automatizzato", nonché l'"affidabilità e verificabilità dei dati e dei percorsi valutativi"; sul piano procedurale, si pone in luce l'importanza dell'informazione in punto di salute e sicurezza, possibilità di valutazione e ricadute occupazionali, richiedendosi all'azienda di indicare preventivamente alle Rsa (o, in loro mancanza, alle organizzazioni sindacali territoriali stipulanti l'accordo territoriale): "tipologia di innovazione, nelle sue componenti tecniche essenziali e nel suo previsto impatto sul ciclo produttivo e sull'assetto organizzativo preesistente; in caso di intelligenze artificiali: i) il tipo di dati che le medesime siano chiamate a raccogliere; ii) i tempi e i modi di conservazione dei dati; iii) le finalità della raccolta dei medesimi".

Rispetto al tema della gestione algoritmica, particolare risonanza mediatica ha riscontrato il Protocollo di intesa del 28 novembre 2023 sottoscritto da Assoespressi e le aziende appaltatrici dell'Emilia-Romagna che si occupano delle consegne 'ultimo miglio' nella filiera Amazon, a cura dei driver: al suo interno, Assoespressi si è resa disponibile a effettuare incontri specifici con le aziende ed eventualmente con le organizzazioni sindacali, su richiesta delle stesse, per verificare situazioni di carichi di lavoro eccessivi con particolare riferimento a singole rotte o eventualmente su zone particolari. Sempre in tale sede, si è prevista la promozione a livello aziendale di campagne di sensibilizzazione dei dipendenti per un utilizzo più corretto del device e si è evidenziata, sul tema, la piena rilevanza del ruolo del RLS.

Numerosi gli interventi che si sono registrati lo scorso anno.

Il Protocollo Saipem del febbraio 2024, sottoscritto ai sensi dell'art. 4 St. Lav., ha riguardato l'introduzione e disciplina di *smart cameras* nel cantiere di Argo/Cassiopea e, quindi, la sperimentazione di un sistema di IA nel campo della sicurezza sui posti di lavoro. A tal fine, è stato coinvolto anche l'Organismo paritetico nazionale di HSE (Salute e sicurezza), previsto dal Contratto collettivo nazionale di lavoro.

Spostando lo sguardo sull'ambito delle piattaforme, può essere significativo ricordare l'Accordo stipulato il 19 febbraio 2024 da Assogrocery con Cgil, Cisl e Uil per la disciplina del lavoro su piattaforma digitale degli shopper. All'art. 5, non solo si dispone un obbligo informativo alle rappresentanze sui processi anche parzialmente automatizzati, ma si prevede anche che a tali informazioni segua la possibilità di un confronto.

L'Accordo di rinnovo per le Banche di Credito cooperativo – Casse Rurali e artigiane, del 9 luglio 2024 – oltre a richiamare esplicitamente la *Joint declaration* sulla IA delle Parti sociali europee di settore del maggio 2024 – ha stabilito che l'individuazione di nuove mansioni e figure professionali sia oggetto di soluzioni condivise nel contesto di un Organismo nazionale bilaterale e paritetico sull'impatto delle nuove tecnologie/digitalizzazione.

L'Accordo integrativo aziendale di GlaxoSmithKline SpA sottoscritto in data 26 luglio 2024 con la RSU Femca, Uiltec e Filctem per il quadriennio 2024-2027 prevede, rispetto alla adozione di nuove tecnologie, tra cui quelle di IA, la "condivisione di iniziative efficaci (…) anche attraverso momenti informativi e formativi dedicati alla RSU e alla popolazione aziendale"; in particolare si stabilisce che la parti condividano "in

forma preventiva e strutturata (...) percorsi finanziati attraverso lo strumento di Fondimpresa estesi al maggior numero possibile di lavoratori. A tal fine, anche con riferimento all'evoluzione tecnologica e digitale degli strumenti di lavoro, le Parti intendono investire in percorsi formativi dedicati".

L'Accordo del 23 settembre 2024 tra la Sanofi S.p.a. e le strutture territoriali di Femca-Cisl, Filctem-Cgil e Uiltec-Uil, in assistenza della RSU, ha previsto un *Patto per il digitale e l'Intelligenza artificiale*, con impegno a verificarne gli effetti. Il Patto è stato redatto dall'Osservatorio digitale bilaterale, costituito dalle parti nel 2022 ai sensi di quanto previsto dal CCNL Industria chimica. In tale sede, sono enunciati alcuni principi generali guida, a garanzia di un approccio alle tecnologie basato sulla centralità del ruolo umano, il principio di accountability, la protezione dei dati, il rispetto per l'integrità umana, l'equità e inclusione, il benessere comune e la sostenibilità ambientale, nonché la salvaguardia della trasparenza, la promozione del coinvolgimento, lo sviluppo della formazione.

L'Accordo di percorso sulla trasformazione del Gruppo Intesa San Paolo, del 23 ottobre 2024, impegna azienda e organizzazioni sindacali a un confronto costante per monitorare gli effetti per tutto il Gruppo, comprese le ricadute dirette sulla rete filiali, del passaggio alla nuova piattaforma digitale Isytech e dello sviluppo dell'IA, anche attraverso l'individuazione di nuovi mestieri, per i quali le parti si confronteranno, nell'ambito della contrattazione di secondo livello, per definire possibili percorsi di sviluppo professionale e delle competenze. A tal fine si prevede l'attivazione del Comitato trasformazione digitale, che si riunirà indicativamente ogni due mesi e avrà il presidio per tutte le tematiche della Banca digitale e della trasformazione digitale del Gruppo.

Più di recente, il CCNL del settore elettrico dell'11 febbraio 2025 dispone, da un lato che l'Osservatorio bilaterale di settore costituisca la sede anche per esaminare la tematica dell'applicazione dell'AI Act e dell'impatto dell'IA sul lavoro, dall'altro che a livello aziendale siano previsti incontri preventivi di confronto e consultazione tra le azienda e le Organizzazioni sindacali – previa formazione delle parti – per un coinvolgimento nell'analisi e verifica degli esiti dell'uso dell'IA.

Ancora, l''Accordo ponte' del 7 marzo 2025 tra le Segreterie nazionali e territoriali Slc Cgil, Fistel Cisl, Uilcom Uil, unitamente alle RSU/RSA, e l'azienda Konecta include un patto di sistema sulla gestione etica degli strumenti di IA. In particolare, in merito agli strumenti di IA atti a supportare i lavoratori durante lo svolgimento della prestazione lavorativa, sono previste garanzie specifiche per l'utilizzo dello strumento Agent Assist, tra cui l'istituzione di un Comitato paritetico, 8 componenti aziendali e 8 componenti sindacali, che monitori il funzionamento dello strumento e verifichi il rispetto di quanto sancito nell'accordo.

Il 19 marzo 2025 è stato sottoscritto un accordo tra la società di assicurazioni CNP Vita Assicura e le RSA Fisac-Cgil, Snfia e Uilca-Uil, per regolare l'adozione dell'IA nei processi aziendali, con la tutela dei diritti dei dipendenti e promozione al contempo dell'innovazione responsabile. In particolare, l'accordo prevede che "per tutti coloro che dovessero essere impattati nelle loro attività lavorative dall'introduzione delle suddette nuove tecnologie, l'azienda garantisce corsi di formazione ad hoc che consentano una riqualificazione e, ove necessario e/o compatibile con le esigenze aziendali, una ricollocazione in altra mansione, coerente con la professionalità, l'inquadramento e l'esperienza acquisita". Al contempo, si vieta l'utilizzo dell'IA "nei processi industriali aziendali se l'impiego di questa nuova tecnologia genera ricadute occupazionali, intendendosi per tali la perdita del posto di lavoro, sul personale assunto a tempo indeterminato alla data del presente accordo".

Se il 31 marzo 2025 è stato indetto uno sciopero da parte delle sigle sindacali del settore TLC per lo stallo nella rinegoziazione del CCNL, pare utile ricordare come proprio all'interno della piattaforma sindacale presentata per tale rinnovo si trovi uno dei profili più interessanti e innovativi nella prospettiva d'analisi che ci occupa, ossia l'istituzione della figura del delegato per la gestione dei dati (a calco, quindi, della figura del RLS prevista per l'ambito della salute e sicurezza sul lavoro[28]).

Il 15 aprile 2025 è stato sottoscritto l'accordo di rinnovo del CCNL per i settori chimico e farmaceutico, valido per il triennio 2025-2028, all'interno del quale trovano sede delle linee guida dedicate all'utilizzo dell'IA nell'Organizzazione aziendale e un Patto sociale per le competenze attento ai profili della transizione digitale ed ecologica. Sul primo versante, viene data particolare rilevanza alla mappatura e analisi di tutti i rischi, da effettuare prima dell'introduzione dei sistemi, nonché

28 La prospettiva di un rappresentante specializzato in materia era stata avanzata in dottrina da Ingrao (2023).

alla trasparenza e al coinvolgimento dei lavoratori e dei loro rappresentanti: adeguata informazione e formazione ai lavoratori coinvolti; informazione preventiva alla RSU sull'introduzione di progetti di IA; coinvolgimento di lavoratori e RSU per la comprensione di scopi, procedure, rischi e benefici.

Considerazioni conclusive

"Riusciamo a vedere solo per un breve tratto davanti a noi, ma già lì vediamo moltissime cose da fare". Con queste parole Alan Turing chiudeva il saggio *Computing Machinery and Intelligence* comparso nel 1950 sulla rivista *Mind*. L'IA è una tecnologia da organizzare e regolare. In queste pagine, da due angolature, abbiamo provato a mettere in fila alcuni contenuti per capire di che cosa stiamo parlando: un primo passaggio che riteniamo necessario. In termini analitici, di progettazione organizzativa e di regolazione davanti a noi "vediamo moltissime cose da fare". È vero che cogliamo solo un breve tratto del percorso, ma osando, sul filo dell'ossimoro, dovremmo sviluppare una lettura sul breve periodo con una prospettiva che guarda lontano. Il lavoro sta mutando rapidamente, il mondo del lavoro sta raggiungendo un livello di eterogeneità interna forse mai visto in epoche precedenti. Ai lavori segnati da processi organizzativi e contenuti operativi piuttosto *tradizionali*, vediamo affiancarsi *nuovi scenari*, ormai del tutto maturi, dentro i quali l'IA sta giocando un ruolo di estrema rilevanza. All'interno di questo perimetro dobbiamo generare dialogo fra le discipline che studiano il lavoro, così come fra gli attori che regolano il lavoro. Nei prossimi anni sarà centrale creare competenze e spazi per discutere quanto sta avvenendo attorno al lavoro, per comprendere e governare un processo di *mutamento antropologico* che passa *anche* attraverso il lavoro.

Bibliografia

Adler D. (2024), *Il duplice enigma. Intelligenza Artificiale e intelligenza umana*, Torino, Einaudi

Aimo M. (2024), Trasparenza algoritmica nel lavoro su piattaforma: quali spazi per i diritti collettivi nella proposta di Direttiva in discussione?, *Lavoro Diritti Europa*, n.2, 6 maggio

Bassini M. (2024), Oggetto, campo di applicazione e ambito territoriale, in Pollicino O., Donati F., Finocchiaro G., Paolucci F. (a cura di), *La disciplina dell'Intelligenza Artificiale*, Milano, Giuffré, pp.113 ss.

Bengio Y. (2016), Macchine che imparano, *Le Scienze*, *Speciale I Quaderni: Intelligenza Artificiale*, n.576

Bradford A. (2020), *The Brussels Effect: How the European Union Rules the World*, Oxford, Oxford University Press

Buoso S. (2020), *Principio di prevenzione e sicurezza sul lavoro*, Torino, Giappichelli

Butera F., De Michelis G. (2024), *Intelligenza Artificiale e lavoro, una rivoluzione governabile*, Venezia, Marsilio

Canal T., Gosetti G., Luppi M. (2024), Qualità del lavoro e digitalizzazione. Riflessioni aperte sul caso Italiano, *Sinappsi*, XIV, n.2, pp.66-92

Castel R. (2019), *La metamorfosi della questione sociale. Una cronaca del salariato*, Sesto San Giovanni (MI), Mimesis

Castel R. (2004), *L'insicurezza sociale. Cosa vuol dire essere protetti*, Torino, Einaudi

Ciucciovino S. (2024), Risorse umane e Intelligenza Artificiale alla luce del regolamento (UE) 2024/1689, tra norme legali, etica e codici di condotta, *Diritto delle Relazioni Industriali*, n.3, pp.573-614

Commissione europea (2025), *Analysis of EU AI Office stakeholder consultations: defining AI systems and prohibited applications. Final study report*, Luxembourg, Publications Office of the European Union

Crawford K. (2021), *Né intelligente né artificiale. Il lato oscuro dell'IA*, Bologna, il Mulino

Crozier M., Friedberg H. (1978), *Attore sociale e sistema. Sociologia dell'azione organizzata*, Milano, Etas

De Masi D. (2018), *Il lavoro nel XXI secolo*, Torino, Einaudi

Dejours C. (2020), *Lavoro vivo*, Sesto San Giovanni (MI), Mimesis

Domingos P. (2018), Le nostre frontiere digitali, *Le Scienze*, n.603, pp.84-89

Esposito E. (2022), *Comunicazione artificiale. Come gli algoritmi producono intelligenza sociale*, Milano, Bocconi University Press

European Commission (2025), *Annex to the Communication to the Commission - Approval of the content of the draft Communication from the Commission - Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (*AI Act*)*, Brussels, 6.2.2025, C(2025) 924 final) ANNEX

Faioli M. (2025), Adeguatezza ex art. 2086 c.c. e obbligo di introdurre tecnologia avanzata per mitigare i rischi da lavoro, *federalismi.it*, n.6, pp.143-157

Falletta P., Marsano A. (2024), Intelligenza Artificiale e protezione dei dati personali: il rapporto tra Regolamento europeo sull'Intelligenza Artificiale e GDPR, *Rivista italiana di informatica e diritto*, n.1, pp.119-137

Feiler L., König B. (2025), Article 3. Definitions, in Pehlivan C.N., Forgó N., Valcke P. (eds.), *The EU Artificial Intelligence (AI) Act. A Commentary*, Alphen aan den Rijn, Wolters Kluwer, p.57

Finocchiaro G. (2024), *Intelligenza Artificiale. Quali regole?*, Bologna, il Mulino

Floridi L. (2022), *Etica dell'Intelligenza Artificiale. Sviluppi, opportunità, sfide*, Milano, Raffaello Cortina Editore

Ford M. (2022), *Il dominio dei robot. Come l'Intelligenza Artificiale rivoluzionerà l'economia, la politica e la nostra vita*, Milano, il Saggiatore

Gaudio G. (2024), Valutazione di impatto e management algoritmico, *Rivista giuridica del lavoro e della previdenza sociale*, n.4, p.538 ss.

Gigerenzer G. (2023), *Perché l'intelligenza umana batte ancora gli algoritmi*, Milano, Raffaello Cortina Editore

Gosetti G. (2024), *Percorsi di sociologia del lavoro. Da Taylor alla società dei lavori*, Milano, Franco Angeli

Gosetti G. (2022), *La qualità della vita lavorativa. Lineamenti per uno studio sociologico*, Milano, Franco Angeli

Gosetti G., Peruzzi M., Calabrese B. (2024), *Digitalizzazione e lavoro. Elementi per un'analisi giuridica e sociologica*, Torino, Giappichelli

Harari Y.N. (2024), *Nexus. Breve storia delle reti di informazione*, Milano, Bompiani

Honneth A. (2025), *Il lavoratore sovrano. Lavoro e cittadinanza democratica*, Bologna, il Mulino

Iannuzzi A. (2025), La definizione di Intelligenza Artificiale fra Regolamento europeo e norma tecnica internazionale, in Pizzetti F. (a cura di), *La regolazione europea dell'Intelligenza Artificiale nella società digitale*, Torino, Giappichelli, pp.11 ss.

Imberti L. (2023), Intelligenza Artificiale e sindacato. Chi controlla i controllori artificiali?, *federalismi.it*, n.29, pp.192-201

Ingrao A. (2023), Controllo a distanza e privacy del lavoratore alla luce dei principi di finalità e proporzionalità della sorveglianza, *Labour & Law Issues*, 9, n.1, pp.101-121

Italiano G.F., Civitarese Matteucci S., Perrucci A. (2022), L'Intelligenza Artificiale: dalla ricerca scientifica alle sue applicazioni. Una introduzione di contesto, in Pajno A., Donati F., Perrucci A. (a cura di), *Intelligenza Artificiale e diritto: una rivoluzione? Diritti fondamentali, dati personali e regolazione*, vol. I, Bologna, il Mulino, pp.43 ss.

Kaplan J. (2024), *Generative A.I. Conoscere capire e usare l'Intelligenza Artificiale generativa*, Roma, Luiss University Press

Lai M. (2025), Brevi note in tema di Intelligenza Artificiale e salute e sicurezza del lavoro, *Lavoro Diritti Europa*, n.1, 9 marzo

Lazzari C., Pascucci P. (2024), Sistemi di IA, salute e sicurezza sul lavoro: una sfida al modello di prevenzione aziendale, fra responsabilità e opportunità, *Rivista giuridica del lavoro e della previdenza sociale*, n.4, Parte I, pp.587 ss.

Loi P. (2001), *La sicurezza. Diritto e fondamento dei diritti nel rapporto di lavoro*, Torino, Giappichelli

Mantelero A. (2024), The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template, *Computer Law and Security Review*, n.54, pp.9 ss.

Mantelero A., Peruzzi M. (2024), L'AI Act e la gestione del rischio nel sistema integrato delle fonti, *Rivista giuridica del lavoro e della previdenza sociale*, n.4, Parte I, pp.517 ss.

Marconi M. (2025), L'intelligenza di Turing, in Turing A.M., *Macchine calcolatrici e intelligenza*, Torino, Einaudi

McQuaid J. (2022), Spiare le emozioni, *Le Scienze*, n.642, pp.42-49

Mézard M. (2024), L'intelligenza artificiale a supporto della scienza, *eco*, n.5, pp.40-43

Mitchell M. (2022), *L'Intelligenza Artificiale. Una guida per esseri umani pensanti*, Torino, Einaudi

Novella M. (2022), Impresa, in Novella M., Tullini P. (a cura di), *Lavoro digitale*, Torino, Giappichelli, pp.5 ss.

Orofino M. (2025), La regolazione dei modelli di IA per finalità generali, in Pizzetti F. (a cura di), *La regolazione europea dell'Intelligenza Artificiale nella società digitale*, Torino, Giappichelli, pp.95 ss.

Paliokaité A., Christenko A., Aloisi A., Riedel L., Ratti M., De Stefano V. (2025), *Study exploring the context, challenges, opportunities, and trends in algorithmic management*, Luxembourg, Publications Office of the European Union

Perilli L. (2025), *Coscienza artificiale. Come le macchine pensano e trasformano l'esperienza umana*, Milano, il Saggiatore

Peruzzi M. (2024), IA e obblighi datoriali di tutela del lavoratore: necessità e declinazioni dell'approccio risk-based, *federalismi.it*, n.30, pp.225-249

Peruzzi M. (2023), *Intelligenza Artificiale e lavoro. Uno studio su poteri datoriali e tecniche di tutela*, Torino, Giappichelli

Riede L., Talhoff O. (2025), Article 25. Responsibilities Along the AI Value Chain, in Pehlivan C. N., Forgó N., Valcke P. (eds.), *The EU Artificial Intelligence (AI) Act. A Commentary*, Alphen aan den Rijn, Wolters Kluwer, pp.516 ss.

Riva G. (2025), *Io, noi, loro. Le relazioni nell'era dei social e dell'IA*, Bologna, il Mulino

Runciman D. (2024), *Affidarsi. Come abbiamo ceduto il controllo della nostra vita a imprese, Stati e intelligenze artificiali*, Torino, Einaudi

Sareddy M.R., Farhan M. (2025), Automatic Resume Screening Approach Based on AI and Ethereum Blockchain for Human Resource Management Using Gaaru, *SN Computer Science*, 6, n.4 <DOI:10.1007/s42979-025-03787-8>

Sartor G., De Gregorio G., Muto G. (2024), Large Language Models, Intelligenza Artificiale generativa e Artificial Intelligence Act, in Pollicino O., Donati F., Finocchiaro G., Paolucci F. (a cura di), *La disciplina dell'Intelligenza Artificiale*, Milano, Giuffré, pp.259 ss.

Schmidl M., Rohner A. (2025), Article 70. Designation of National Competent Authorities and Single Point of Contact, in Pehlivan C.N., Forgó N., Valcke P. (a cura di), *The EU Artificial Intelligence (AI) Act. A Commentary*, Alphen aan den Rijn, Wolters Kluwer, pp.1019 ss.

Sennett R. (2000), *L'uomo flessibile. Le conseguenze del nuovo capitalismo sulla vita personale*, Milano, Feltrinelli

Spitzer M. (2024), *Intelligenza Artificiale. Opportunità e rischi di una rivoluzione tecnologica che sta cambiando il mondo*, Milano, Corbaccio

Tebano L. (2020), *Lavoro, potere direttivo e trasformazioni organizzative*, Napoli, Editoriale scientifica

Tönnies F. (1887), *Gemeinschaft und Gesellschaft. Abhandlung des Communismus und des Socialismus als empirischer Culturfomen*, Leipzig, Fues

Treu T. (2025), *Il controllo umano delle tecnologie: regole e procedure*, WP C.S.D.L.E. "Massimo D'Antona".IT n.492, Catania, Centre for the Study of European Labour Law "MASSIMO D'ANTONA", University of Catania

Treu T. (2024), *Intelligenza Artificiale (IA): integrazione o sostituzione del lavoro umano?*, WP C.S.D.L.E. "Massimo D'Antona". IT n.487, Catania, Centre for the Study of European Labour Law "MASSIMO D'ANTONA", University of Catania

Tullini P. (2021), La questione del potere nell'impresa. Una retrospettiva lunga mezzo secolo, *Lavoro e diritto*, n.3-4, pp.442 ss.

Turing A. (1950), Computing Machinery and Intelligence, *Mind*, LIX, n.236, pp.433-460

Varanini F. (2024), *Splendori e miserie delle intelligenze artificiali. Alla luce dell'esperienza umana*, Milano, Guerini e associati

Voigt P., Hullen N. (2024), *The EU AI Act. Answers to Frequently Asked Questions*, Berlin, Heidelberg, Springer

Zuboff S. (2019), *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, Roma, Luiss University Presse

Giorgio Gosetti
giorgio.gosetti@univr.it
Insegna Sociologia del lavoro presso il Dipartimento di Scienze umane dell'Università degli Studi di Verona ed è Responsabile scientifico del Centro di ricerca RE-WOrk (REsearching for REmaking Work and Organizing). Fra le sue pubblicazioni recenti si segnalano: (con Peruzzi M., Calabrese B.), *Digitalizzazione e lavoro. Elementi per un'analisi giuridica e sociologica*, Giappichelli, 2024; *Percorsi di sociologia del lavoro. Da Taylor alla società dei lavori*, Franco Angeli, 2024; *La qualità della vita lavorativa. Lineamenti per uno studio sociologico*, Franco Angeli, 2022.

Marco Peruzzi
marco.peruzzi@univr.it
Insegna Diritto del lavoro presso il Dipartimento di Scienze giuridiche dell'Università degli Studi di Verona. Fra le sue pubblicazioni recenti si segnalano: IA e obblighi datoriali di tutela del lavoratore: necessità e declinazioni dell'approccio risk-based, *Federalismi.it*, 2024; Gestione algoritmica del lavoro, protezione dei dati personali e tutela collettiva, *Lavoro e diritto*, n.2, 2024; *Intelligenza artificiale e lavoro*, Giappichelli, 2023.

Adozione di una metodologia di valutazione di conformità etica di sistemi intelligenti governativi

Gaetano Riccio
Università degli Studi di Napoli Federico II

L'integrazione dell'Intelligenza artificiale Generativa (IAG) nella Pubblica amministrazione offre opportunità per ottimizzare decisioni, servizi e interazioni con i cittadini, ma solleva complessità normative ed etiche. Lo studio propone un framework che armonizza AI Act, GDPR e standard internazionali, adattati al settore pubblico. Governance multilivello, gestione del rischio e supervisione umana emergono come strumenti essenziali per mitigare *bias*, garantire trasparenza e tutelare diritti. L'analisi di casi europei mostra che una governance etica e robusta rafforza compliance, sicurezza e fiducia, favorendo un valore pubblico sostenibile.

The integration of Generative Artificial Intelligence (GAI) into Public Administration offers opportunities to optimize decision-making, services, and citizen interactions, while raising regulatory and ethical challenges. This study proposes a framework harmonizing the AI Act, GDPR, and international standards, tailored to the public sector. Multilevel governance, risk management, and human oversight emerge as key tools to mitigate bias, ensure transparency, and safeguard rights. Analysis of European cases shows that ethical, robust governance strengthens compliance, security, and trust, fostering sustainable public value.

DOI: 10.53223/Sinappsi_2025-02-11

Citazione
Riccio G. (2025), Adozione di una metodologia di valutazione di conformità etica di sistemi intelligenti governativi, *Sinappsi*, XV, n.2, pp.156-170

Parole chiave
Governance
Intelligenza artificiale
Pubblica amministrazione

Keywords
Governance
Artificial intelligence
Public administration

Introduzione: il contesto di adozione dell'IA generativa nella Pubblica amministrazione

L'evoluzione dell'Intelligenza artificiale ha raggiunto una fase cruciale con l'avvento dei sistemi generativi, che non si limitano ad analizzare dati esistenti ma possono creare autonomamente nuovi contenuti testuali, visivi e multimediali. Questi sistemi, basati su architetture avanzate come i *Large Language Models* (LLM) e le reti generative avversarie (GAN), offrono alla Pubblica amministrazione (PA) potenzialità trasformative senza precedenti. A differenza dei sistemi di IA tradizionali, focalizzati su compiti specifici e predeterminati, l'IA generativa permette di automatizzare la produzione di documenti amministrativi, fornire assistenza personalizzata ai cittadini attraverso chatbot avanzati, generare analisi predittive complesse e supportare processi decisionali con informazioni contestualizzate. Questa rivoluzione tecnologica si inserisce nel più ampio processo di digitalizzazione

della PA, accelerato dalla pandemia di Covid-19 e dagli investimenti previsti dal Piano nazionale di ripresa e resilienza (PNRR).

La questione di ricerca che questo articolo intende affrontare riguarda le modalità con cui le amministrazioni pubbliche possano implementare sistemi di IA generativa garantendo contemporaneamente la conformità normativa, la tutela dei diritti fondamentali e l'efficienza amministrativa. La tesi sostenuta è che un'implementazione efficace richieda un framework integrato che coordini aspetti normativi, organizzativi, tecnici ed etici in un approccio olistico, superando la frammentazione che caratterizza molte iniziative attuali. Secondo Misuraca *et al.* (2020), infatti, le iniziative di IA nella PA europea tendono a svilupparsi in maniera non coordinata, creando disallineamenti tra innovazione tecnologica e framework regolatori.

Il contesto normativo europeo si è evoluto significativamente con l'adozione dell'AI Act[1], che impone requisiti specifici per i sistemi di IA ad alto rischio, molti dei quali trovano applicazione nell'ambito pubblico. Tale Regolamento si aggiunge al GDPR, creando un complesso ecosistema normativo che le PA devono navigare nell'implementazione di soluzioni di IA generativa. Parallelamente, standard internazionali come il NIST AI Risk Management Framework (AI RMF) e l'ISO 42001 forniscono linee guida complementari sulla gestione dei rischi e sulla governance dei sistemi di IA. Questa stratificazione normativa richiede un approccio integrato che armonizzi i diversi requisiti in un framework operativo coerente.

La letteratura scientifica ha evidenziato diversi gap nella ricerca sull'implementazione dell'IA generativa nella PA. Wirtz *et al.* (2022) identificano una carenza di studi empirici sugli impatti organizzativi dell'IA nel settore pubblico, mentre Sun e Medaglia (2019) sottolineano la mancanza di framework specifici per la valutazione del rischio dei sistemi generativi. Un'analisi condotta da Cugurullo (2021) evidenzia inoltre come molte amministrazioni pubbliche implementino soluzioni di IA senza adeguate valutazioni preliminari di impatto sui diritti fondamentali, creando potenziali vulnerabilità legali e sociali. Il presente articolo intende contribuire a colmare questo gap, analizzando un framework integrato per l'implementazione dell'IA generativa nella PA che bilanci innovazione tecnologica e compliance normativa.

1. Quadro teorico e letteratura di riferimento

L'Intelligenza artificiale generativa rappresenta un sottoinsieme dell'IA caratterizzato dalla capacità di creare contenuti originali piuttosto che limitarsi ad analizzare dati esistenti. Secondo la definizione di Radford *et al.* (2019), l'IAG comprende modelli computazionali progettati per generare contenuti nuovi che imitano caratteristiche statistiche dei dati di addestramento. Questi sistemi, basati principalmente su architetture *transformer* e *deep learning*, hanno conosciuto un'accelerazione esponenziale negli ultimi anni, culminando con modelli come GPT (Generative Pre-trained Transformer), DALL-E e Midjourney, che hanno democratizzato l'accesso a potenti strumenti generativi. La loro adozione nella PA introduce potenzialità trasformative, ma solleva anche questioni uniche rispetto ai sistemi di IA tradizionali, particolarmente in termini di affidabilità, controllo dei contenuti e responsabilità giuridica.

La letteratura accademica sull'implementazione dell'IA nel settore pubblico ha iniziato a espandersi significativamente, ma presenta ancora lacune rilevanti quando si focalizza specificamente sui sistemi generativi. Wirtz e Müller (2023) hanno condotto una revisione sistematica identificando 135 articoli accademici sull'IA nella PA pubblicati tra 2000 e 2022, rilevando come solo il 12% affronti questioni di governance e il 7% aspetti etici. Ancora più esigua è la letteratura specifica sull'IA generativa nella PA, con studi pionieristici come quello di Criado e Gil-Garcia (2019) che esplorano le potenzialità delle chatbot avanzate nell'erogazione di servizi pubblici, ma sottolineano la carenza di framework implementativi completi.

Il quadro teorico per l'analisi dell'IAG nella PA può essere articolato attraverso tre prospettive complementari. La prima è la teoria del valore pubblico (Moore 1995; Twizeyimana e Andersson

1 Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (Regolamento sull'Intelligenza artificiale).

2019), che valuta l'impatto dell'IAG in termini di miglioramento dei servizi, efficienza operativa e fiducia istituzionale. La seconda prospettiva è quella dell'etica digitale (Floridi e Cowls 2019), che esamina le implicazioni morali dell'automazione decisionale nella sfera pubblica. La terza è la teoria dell'accettazione tecnologica nel settore pubblico (Venkatesh *et al.* 2016), che analizza i fattori organizzativi e culturali che influenzano l'adozione di tecnologie avanzate nelle burocrazie pubbliche.

Una revisione della letteratura evidenzia tre principali gap di ricerca. Innanzitutto, vi è una carenza di studi empirici sull'implementazione dell'IAG nelle PA, con poche analisi dei fattori di successo e delle barriere organizzative. Van der Voort *et al.* (2019) sottolineano come la maggior parte degli studi si concentri su aspetti tecnici piuttosto che su dimensioni implementative e organizzative. In secondo luogo, mancano framework integrati che armonizzino requisiti normativi, considerazioni etiche e necessità operative. Misuraca *et al.* (2020) evidenziano la frammentazione degli approcci, con modelli che tendono a focalizzarsi su singoli aspetti (conformità normativa, performance tecnica, accettazione organizzativa) senza una visione olistica. Infine, vi è una carenza di ricerche sulle specificità dell'IAG rispetto all'IA tradizionale nel contesto pubblico, con poca attenzione alle peculiari sfide poste dai sistemi generativi in termini di controllo dei contenuti, responsabilità giuridica e integrazione nei processi decisionali pubblici.

Come sottolineato da Veale e Borgesius (2021), l'efficace governance dell'IA nel settore pubblico richiede l'integrazione di molteplici framework regolatori in un approccio coerente e operativamente sostenibile, una sfida che questo articolo intende affrontare attraverso l'analisi di un modello implementativo specifico per l'IAG.

2. Metodologia della ricerca

La presente ricerca adotta un approccio metodologico misto, combinando analisi documentale, valutazione di framework e studio di casi. Questa triangolazione metodologica permette di esaminare il fenomeno dell'implementazione dell'IAG nella PA da prospettive complementari, aumentando la robustezza e la validità dei risultati. L'approccio si articola in tre fasi principali, ciascuna caratterizzata da specifiche tecniche di raccolta e analisi dei dati.

La prima fase ha comportato un'analisi sistematica della letteratura accademica sull'implementazione dell'IA nella PA, con particolare attenzione ai sistemi generativi. Seguendo la metodologia PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) proposta da Moher *et al.* (2009), sono stati selezionati 78 articoli accademici da database come Scopus, Web of Science e Google Scholar, utilizzando combinazioni di parole chiave come 'Generative Aì, 'public administration', 'implementation framework' e 'regulatory compliance'. Gli articoli sono stati sottoposti a screening e classificati secondo dimensioni analitiche predefinite (focus tecnologico, aspetti normativi, considerazioni etiche, barriere implementative). Questa fase ha permesso di identificare i principali gap nella letteratura e le dimensioni cruciali per un framework implementativo efficace.

La seconda fase ha comportato l'analisi approfondita del Framework di Intelligenza artificiale per la Pubblica amministrazione presentato nell'Allegato a questo contributo. L'analisi è stata condotta utilizzando una griglia valutativa sviluppata integrando il modello di Bannister e Connolly (2014) per la valutazione degli strumenti di e-government e l'approccio di Janowski *et al.* (2022) per l'analisi dei framework di governance dell'IA. La griglia ha permesso di valutare il framework secondo quattro dimensioni principali: completezza normativa, usabilità operativa, adeguatezza contestuale e solidità metodologica. Per ciascuna dimensione sono stati definiti indicatori specifici, valutati su una scala Likert a cinque punti.

La raccolta dati ha comportato l'analisi di documentazione progettuale, interviste semi-strutturate con 12 stakeholder chiave (dirigenti pubblici, tecnici, responsabili per la protezione dei dati) e l'osservazione diretta dei sistemi implementati. L'analisi dei casi è stata condotta utilizzando il framework sviluppato da Yin (2018) per gli studi di caso comparativi, con particolare attenzione all'identificazione di pattern ricorrenti, fattori abilitanti e barriere implementative.

I limiti metodologici includono la rappresentatività limitata dei casi studiati, che non possono catturare l'intera varietà di applicazioni dell'IAG nella PA, e la rapidità dell'evoluzione tecnologica e normativa, che può rendere alcuni risultati rapidamente obsoleti. Inoltre, essendo l'implementazione dell'IAG nella PA un fenomeno relativamente recente, non è

stato possibile valutare gli impatti a lungo termine delle soluzioni analizzate. Nonostante questi limiti, l'approccio metodologico adottato fornisce una base solida per l'analisi del framework implementativo e l'identificazione di linee guida operative per l'adozione dell'IAG nel settore pubblico.

3. Il Framework integrato per l'IA generativa nella PA

Il Framework di Intelligenza artificiale per la Pubblica amministrazione sviluppato e analizzato in questo studio rappresenta un approccio strutturato all'implementazione di sistemi di IA, inclusi quelli generativi, nel contesto del settore pubblico. La sua caratteristica distintiva risiede nell'integrazione di requisiti normativi, considerazioni etiche e necessità operative in un modello coerente e applicabile. Il framework si articola in quattro dimensioni principali: governance strategica, valutazione del rischio, sicurezza cibernetica e formazione, ciascuna adattata alle specificità dell'IAG e alle peculiarità del contesto pubblico.

La governance strategica costituisce il pilastro fondamentale del framework, ponendo le basi per un'implementazione allineata sia agli obiettivi istituzionali che ai requisiti normativi. Il modello propone una struttura di governance multilivello che comprende un Comitato etico-tecnologico interdisciplinare, un Registro centralizzato dei sistemi IA e un Piano triennale di adozione. Come evidenziato da Janssen e van der Voort (2016), un'efficace governance dell'IA richiede strutture istituzionali che bilancino competenze tecniche, giuridiche ed etiche in un approccio adattivo, principio che il framework implementa attraverso la composizione multidisciplinare del Comitato. Particolarmente rilevante per i sistemi generativi è la previsione di un AI Governance Board (AGB) che include non solo dirigenti e tecnici, ma anche esperti legali e rappresentanti etici provenienti da ONG e università, garantendo così un approccio multistakeholder alla governance.

L'allineamento agli standard internazionali costituisce un altro elemento distintivo del framework, che incorpora principi e metodologie dal NIST AI RMF, dall'ISO 42001 e dal MIT AI Risk Repository. Questa integrazione risponde alla raccomandazione di Lewis *et al.* (2022) di armonizzare standard tecnici e normativi in un approccio coerente alla governance dell'IA. Particolarmente significativa per l'IAG è l'adozione del Generative AI Profile del NIST, che introduce parametri specifici di robustezza per sistemi come chatbot e assistenti virtuali, affrontando così le peculiari sfide di controllo qualitativo poste dai sistemi generativi.

La valutazione del rischio rappresenta una componente cruciale del framework, particolarmente rilevante per l'IAG dato il suo carattere potenzialmente imprevedibile. Il modello propone un'innovativa Valutazione integrata che combina la Fundamental Rights Impact Assessment (FRIA) richiesta dall'AI Act con la Data Protection Impact Assessment (DPIA) prevista dal GDPR. Questa integrazione risponde alla necessità, evidenziata da Kaminski e Malgieri (2021), di superare la frammentazione delle valutazioni d'impatto in un approccio olistico alla gestione del rischio. Il framework prevede metodologie specifiche per l'identificazione del *bias* nei dataset, utilizzando strumenti come AI Fairness 360, e protocolli per test di robustezza contro attacchi adversarial, particolarmente rilevanti per i sistemi generativi esposti a tentativi di manipolazione.

Un elemento distintivo del framework è l'enfasi sui meccanismi '*human-in-the-loop*' per garantire supervisione umana e spiegabilità delle decisioni automatizzate. Come sottolineato da Veale e Borgesius (2021), i sistemi generativi richiedono modelli di supervisione adattivi che combinino controllo preventivo e monitoraggio continuo. Il framework implementa questo principio attraverso dashboard di monitoraggio in tempo reale che segnalano anomalie statistiche, soglie di incertezza per l'attivazione di canali umani e meccanismi di spiegabilità 'a gradiente' che forniscono diversi livelli di dettaglio esplicativo a seconda della criticità della decisione.

La sicurezza cibernetica costituisce un altro pilastro del framework, con particolare attenzione ai rischi specifici dei sistemi generativi. Il modello prevede un approccio di 'security by design' con particolare enfasi sulla protezione contro attacchi mirati come il *data poisoning*, particolarmente pericolosi per i sistemi basati su LLM. Significativa è l'introduzione di tecniche di crittografia omomorfica che permettono di elaborare dati sensibili senza esporli in chiaro, rispondendo così alla raccomandazione di Floridi *et al.* (2020) di implementare protezioni tecniche

che garantiscano privacy by design nei sistemi di IA pubblica. Il framework prevede inoltre protocolli specifici per la gestione degli incidenti che integrano i requisiti del GDPR con quelli dell'AI Act, garantendo notifiche tempestive e analisi post-mortem strutturate.

Infine, il framework dedica ampio spazio alla formazione e al cambiamento culturale, riconoscendo che l'implementazione efficace dell'IAG richiede non solo strumenti tecnici ma anche adeguate competenze organizzative. Il modello prevede corsi certificati su *AI Ethics*, simulazioni di caso su scenari reali e toolkit dipartimentali con linee guida specifiche, ad esempio sul prompt engineering per sistemi generativi. Questo approccio risponde alla raccomandazione di Mergel *et al.* (2019) di investire nello sviluppo di competenze trasversali che combinino comprensione tecnologica ed etica applicata.

4. Sfide implementative e fattori critici di successo

L'implementazione di sistemi di Intelligenza artificiale generativa nella Pubblica amministrazione presenta sfide distintive che richiedono strategie mirate. L'analisi dei casi studio e della letteratura ha permesso di identificare cinque aree critiche di sfida: resistenza organizzativa, carenza di competenze, integrazione nei processi esistenti, gestione dell'incertezza algoritmica e sostenibilità economica. Ciascuna di queste dimensioni presenta specificità legate all'IAG e richiede approcci gestionali differenziati.

La resistenza organizzativa emerge come una barriera significativa, manifestandosi in diverse forme: dalla diffidenza verso l'automazione di funzioni tradizionalmente umane al timore di perdita di controllo decisionale. Come evidenziato da Mergel *et al.* (2019**)**, l'introduzione di sistemi di IA nella PA si scontra con culture organizzative consolidate e strutture burocratiche rigide. Nel caso dell'IAG, questa resistenza assume caratteristiche peculiari legate alla percezione di imprevedibilità dei sistemi generativi. L'analisi del caso dell'amministrazione comunale italiana ha evidenziato come la resistenza fosse particolarmente accentuata per funzioni percepite come 'creative' o richiedenti discrezionalità amministrativa. Strategie efficaci di mitigazione hanno incluso approcci di co-design che hanno coinvolto i funzionari nella definizione dei parametri operativi del sistema generativo e la graduale introduzione dell'automazione, mantenendo significativi spazi di controllo umano.

La carenza di competenze rappresenta una sfida particolarmente acuta per l'IAG, che richiede non solo conoscenze tecniche ma anche capacità di valutazione critica degli output generati. Wirtz e Müller (2023) sottolineano come il digital divide interno alle PA costituisca un ostacolo maggiore all'adozione dell'IA rispetto a vincoli tecnologici o economici. I casi analizzati hanno evidenziato la necessità di un triplice livello di competenze: tecniche (comprendere i meccanismi di funzionamento), operative (saper interagire efficacemente con i sistemi) e valutative (saper giudicare criticamente gli output). Il framework affronta questa sfida attraverso programmi formativi multilivello, comunità di pratica tra enti e toolkit operativi specifici per diverse funzioni amministrative.

L'integrazione nei processi esistenti costituisce una sfida cruciale, particolarmente complessa per i sistemi generativi che possono produrre output non standardizzati. L'analisi del caso dell'agenzia governativa spagnola ha evidenziato difficoltà di integrazione della chatbot avanzata con i sistemi di gestione documentale esistenti e con i workflow amministrativi consolidati. Janssen e van der Voort (2016) sottolineano l'importanza di approcci di integrazione adattivi che bilancino innovazione e continuità operativa. Strategie efficaci hanno incluso l'adozione di interfacce standardizzate, lo sviluppo di API dedicate e la riprogettazione incrementale dei processi, mantenendo inizialmente percorsi paralleli (automatizzati e tradizionali) per garantire la continuità operativa durante la transizione.

La gestione dell'incertezza algoritmica rappresenta una sfida distintiva dell'IAG, i cui output possono presentare gradi variabili di accuratezza, rilevanza e conformità normativa. Questa sfida assume particolare rilevanza nel contesto pubblico, dove le decisioni amministrative devono rispettare stringenti requisiti di correttezza, imparzialità e tracciabilità. Il caso del Ministero olandese ha evidenziato come la gestione dell'incertezza algoritmica richieda un'articolata strategia di controllo qualità, combinando filtri preventivi (prompt engineering, controlli preventivi), monitoraggio in tempo reale (*confidence scores*, alert su anomalie) e revisione umana post-generazione. Come sottolineato da Floridi *et al.* (2020), i sistemi

generativi richiedono un continuum di supervisione umana calibrato sul livello di criticità della funzione.

La sostenibilità economica, infine, rappresenta una sfida significativa, considerando i costi di implementazione, mantenimento e aggiornamento dei sistemi generativi. L'analisi dei casi ha evidenziato come i costi iniziali di acquisizione tecnologica rappresentino solo una frazione del costo totale di possesso, con significativi investimenti richiesti per formazione, personalizzazione, integrazione e monitoraggio continuo. Il framework propone modelli di condivisione delle risorse tra amministrazioni, approcci incrementali all'implementazione e metodologie di valutazione costi-benefici che considerino non solo l'efficienza operativa, ma anche dimensioni di valore pubblico più ampie come trasparenza, accessibilità e fiducia istituzionale.

5. Raccomandazioni di policy per l'implementazione efficace

Sulla base dell'analisi del framework e dei gap implementativi identificati, è possibile formulare raccomandazioni di policy strutturate per guidare l'efficace adozione dell'IAG nella Pubblica amministrazione. Queste raccomandazioni sono organizzate su tre livelli complementari: strategico, tattico e operativo, fornendo così una roadmap completa per decisori pubblici e implementatori.

A livello strategico, è essenziale sviluppare una visione nazionale coordinata per l'adozione dell'IAG nella PA, superando l'attuale frammentazione delle iniziative. Questa visione dovrebbe tradursi in una Strategia nazionale per l'IA generativa nella PA, integrata nel più ampio Piano triennale per l'Informatica nella Pubblica amministrazione. La strategia dovrebbe definire priorità di investimento, casi d'uso prioritari e meccanismi di coordinamento inter-istituzionale. Come evidenziato da Misuraca *et al.* (2020), l'assenza di una visione strategica coerente rappresenta il principale ostacolo alla scalabilità delle iniziative di IA nel settore pubblico. Un elemento cruciale di questa strategia dovrebbe essere l'istituzione di un Centro di Competenza nazionale sull'IA generativa per la PA, con il mandato di sviluppare linee guida, fornire supporto tecnico e facilitare la condivisione di esperienze tra amministrazioni. Il centro potrebbe essere istituito presso AgID o altra struttura esistente, con un modello di governance che assicuri rappresentanza delle diverse tipologie di amministrazioni e adeguato coinvolgimento di stakeholder accademici e della società civile.

È inoltre raccomandabile lo sviluppo di un Regulatory Sandbox specifico per l'IAG nella PA, che permetta la sperimentazione controllata di applicazioni innovative in ambienti a ridotto rischio regolatorio. Questo meccanismo, già sperimentato con successo in domini come fintech e health-tech, permetterebbe di testare applicazioni generative in ambiti delimitati prima di una più ampia implementazione, facilitando l'innovazione senza compromettere la tutela dei diritti fondamentali. Come sottolineato da Lodge e Mennicken (2019), i *regulatory sandboxes* creano spazi protetti per l'innovazione responsabile, particolarmente preziosi in domini ad alta regolamentazione come la PA.

A livello tattico, è raccomandabile lo sviluppo di Toolkit implementativi settoriali che operazionalizzino il framework generale per specifici domini applicativi (es. servizi sociali, urbanistica, gestione documentale), fornendo linee guida contestualizzate, template per valutazioni d'impatto e checklist operative. L'approccio '*one size fits all*' risulta infatti inadeguato considerando l'eterogeneità delle funzioni amministrative e dei relativi requisiti. Tali toolkit dovrebbero includere modelli di governance adattati alle dimensioni organizzative dell'ente, semplificando le strutture proposte dal framework per piccole amministrazioni senza rinunciare ai principi fondamentali di supervisione multi-stakeholder e di valutazione di impatto.

È inoltre cruciale lo sviluppo di un AI Evaluation Framework standardizzato che permetta alle PA di valutare sistemi di IAG prima dell'acquisizione, verificandone non solo le performance tecniche, ma anche conformità normativa, spiegabilità e robustezza. Tale framework dovrebbe essere incorporato nelle linee guida per gli appalti pubblici, richiedendo ai fornitori evidenze documentate di conformità. Janssen *et al.* (2020) sottolineano come la fase di procurement rappresenti un'opportunità cruciale per incorporare requisiti di responsible AI nelle implementazioni pubbliche, un principio che dovrebbe guidare l'aggiornamento delle procedure di appalto per soluzioni di IAG.

A livello operativo, è raccomandabile l'implementazione di un sistema di Certificazione delle competenze sull'IAG nella PA, che definisca

profili professionali (es. AI Ethics Officer, LLM Prompt Engineer, AI Data Steward) e percorsi formativi certificati. Tale sistema dovrebbe essere integrato nei piani di fabbisogno del personale e nei percorsi di sviluppo professionale, creando incentivi per l'acquisizione di competenze specialistiche. Il sistema di certificazione potrebbe essere sviluppato in collaborazione con università ed enti formativi, garantendo allineamento con competenze richieste dal mercato e specificità del contesto pubblico.

È inoltre essenziale lo sviluppo di una Piattaforma collaborativa che faciliti la condivisione di risorse (modelli, dataset, prompt *libraries*, valutazioni di impatto) tra amministrazioni, riducendo duplicazioni e facilitando l'adozione di soluzioni già testate. Tale piattaforma dovrebbe implementare principi di 'governo come piattaforma' (ÒReilly 2011), facilitando ecosistemi collaborativi piuttosto che approcci top-down. Un esempio di successo è rappresentato dal programma *Developers* Italia, che potrebbe essere esteso con sezioni dedicate a risorse per implementazioni di IAG.

Infine, è raccomandabile l'implementazione di un sistema di monitoraggio e apprendimento continuo che raccolga dati sull'implementazione dell'IAG nelle diverse amministrazioni, identificando pattern di successo, barriere ricorrenti e impatti su diversi stakeholder. Tale sistema dovrebbe adottare metodologie miste (quantitative e qualitative) e prevedere meccanismi di feedback da cittadini e funzionari pubblici. Come sottolineato da de Bruijn *et al.* (2020), i sistemi di performance management per l'IA pubblica devono superare metriche puramente tecniche per incorporare dimensioni di valore pubblico più ampie, un principio che dovrebbe guidare lo sviluppo di framework valutativi per l'IAG nella PA.

Conclusioni e direzioni future di ricerca

L'implementazione dell'Intelligenza artificiale generativa nella Pubblica amministrazione rappresenta una frontiera di innovazione con potenzialità trasformative, ma caratterizzata da sfide complesse che richiedono approcci strutturati e contestualmente sensibili. L'analisi del Framework di Intelligenza artificiale per la Pubblica amministrazione ha evidenziato come un'efficace adozione dell'IAG richieda l'integrazione di molteplici dimensioni: governance strategica, valutazione del rischio, sicurezza cibernetica e formazione, ciascuna adattata alle specificità dei sistemi generativi e alle peculiarità del contesto pubblico. Il framework analizzato offre un approccio olistico che bilancia innovazione tecnologica, compliance normativa e tutela dei diritti fondamentali, rispondendo alla necessità di armonizzare requisiti dell'AI Act, del GDPR e degli standard internazionali in un modello operativamente sostenibile.

L'analisi comparativa tra il framework ideale e le concrete esperienze implementative ha evidenziato significativi gap da colmare, particolarmente nelle aree della governance multi-stakeholder, delle metodologie di valutazione del rischio, dei meccanismi '*human-in-the-loop*' e delle pratiche di sicurezza cibernetica. Questi gap non derivano necessariamente da carenze del framework, ma piuttosto dalle complessità organizzative, tecniche e culturali che caratterizzano i contesti applicativi reali. La loro identificazione offre opportunità di affinamento del framework e sviluppo di strategie implementative più efficaci, in un processo iterativo di apprendimento e adattamento.

Le raccomandazioni di policy proposte offrono una roadmap strutturata per decisori pubblici e implementatori, articolata su tre livelli complementari: strategico (Strategia nazionale, Centro di competenza, Regulatory Sandbox), tattico (Toolkit implementativi settoriali, AI Evaluation Framework) e operativo (Certificazione delle competenze, Piattaforma collaborativa, sistema di monitoraggio). Queste raccomandazioni non rappresentano soluzioni definitive, ma piuttosto orientamenti per un percorso di trasformazione che richiederà continuo adattamento alle evoluzioni tecnologiche, normative e organizzative.

Questa ricerca contribuisce alla letteratura sull'IA nella PA colmando parzialmente i gap identificati, particolarmente riguardo ai framework implementativi specifici per sistemi generativi e all'integrazione di requisiti normativi in modelli operativamente sostenibili. Tuttavia, diversi aspetti richiedono ulteriore approfondimento attraverso ricerche future. Innanzitutto, è necessario sviluppare metodologie più robuste per la valutazione empirica dell'impatto dell'IAG nella PA, superando approcci aneddotici per misurare sistematicamente effetti su efficienza amministrativa, qualità dei servizi e fiducia istituzionale. Come sottolineato da Wirtz e Müller (2023), la carenza di studi empirici longitudinali limita la comprensione degli effetti trasformativi

dell'IA nel settore pubblico, una lacuna che future ricerche dovrebbero colmare.

È inoltre cruciale approfondire le dinamiche di accettazione organizzativa dell'IAG nella PA, identificando fattori culturali, strutturali e individuali che influenzano resistenze e adozione. Particolarmente rilevante sarebbe l'analisi delle interazioni tra sistemi generativi e lavoratori pubblici in contesti reali, esaminando come si riconfigurino ruoli professionali, competenze e identità lavorative. Mergel *et al.* (2019) sottolineano come la trasformazione dei ruoli professionali rappresenti una delle dimensioni meno esplorate dell'adozione dell'IA nella PA, un gap che richiederebbe approcci etnografici e longitudinali.

Un terzo filone di ricerca riguarda i modelli di governance dell'IAG nella PA, con particolare attenzione a meccanismi di accountability democratica e partecipazione civica. Come evidenziato da Veale e Borgesius (2021), l'automazione decisionale nel settore pubblico solleva questioni fondamentali di responsabilità democratica che trascendono aspetti puramente tecnici o legali, richiedendo ricerche interdisciplinari che integrino prospettive giuridiche, politologiche e sociologiche.

Infine, è necessario sviluppare approcci valutativi che misurino l'impatto dell'IAG non solo in termini di efficienza, ma rispetto a valori pubblici più ampi come equità, trasparenza e legittimità. Tali ricerche dovrebbero adottare metodologie miste, combinando analisi quantitative di performance con indagini qualitative su percezioni ed esperienze di diversi stakeholder, in un approccio valutativo multidimensionale e partecipativo.

In conclusione, l'IAG rappresenta per la PA non solo una sfida tecnologica, ma una più ampia opportunità di ripensamento istituzionale, in cui innovazione digitale e valori pubblici fondamentali devono integrarsi in un nuovo paradigma di governance. Come sottolineato da Janowski *et al.* (2022), il successo dell'IA nel settore pubblico si misurerà non solo in termini di efficienza operativa, ma nella capacità di rafforzare legittimità democratica e fiducia istituzionale, una prospettiva che dovrebbe guidare sia l'implementazione pratica, che la ricerca futura in questo campo.

Allegato
Framework di Intelligenza artificiale per la Pubblica amministrazione: integrazione normativa e strumenti operativi

1. Governance strategica e allineamento normativo

1.1 Quadro Giuridico Integrato

L'AI Act e il GDPR costituiscono il fondamento giuridico per l'uso dell'IA nella PA. L'AI Act classifica i sistemi in quattro categorie di rischio (inaccettabile, alto, limitato, minimo), imponendo obblighi specifici per quelli ad alto rischio, come i sistemi di valutazione creditizia o di reclutamento pubblico. Il GDPR, invece, regola il trattamento dei dati personali, richiedendo valutazioni d'impatto (DPIA) e garanzie per i diritti degli interessati.

Per allinearsi a questi requisiti, la PA deve adottare un **modello di governance multilivello**:

- **Comitato etico-tecnologico**: organo interdisciplinare per supervisionare progetti IA, verificando la conformità all'AI Act e ai principi del GDPR. Deve includere rappresentanti legali, esperti di cybersecurity, e delegati per la protezione dei dati (DPO).
- **Registro centralizzato dei sistemi IA**: catalogo aggiornato di tutti i sistemi in uso, con metadata su finalità, rischi, basi giuridiche e cicli di audit. Il registro deve essere accessibile alle autorità di vigilanza e aggiornato trimestralmente.
- **Piano triennale di adozione IA**: allineato alle linee guida AGID e al Piano triennale per l'Informatica 2024-2026, con priorità definite tramite analisi costi-benefici e impatto sociale. Il Piano deve identificare almeno tre progetti pilota a basso rischio (es. ottimizzazione flussi documentali) entro il primo anno.

Checklist operativa:

- Designare un responsabile IA per ogni ente, con competenze in diritto digitale e gestione del rischio.
- Integrare il registro IA con il sistema di gestione documentale (es. Protocollo informatico nazionale).

- Sottoporre il Piano triennale a consultazione pubblica per 60 giorni prima dell'adozione.

1.2 Allineamento agli standard internazionali

Il framework incorpora elementi da:

- **NIST AI RMF**: gestione del rischio tramite categorie di threat (sicurezza, *bias*, privacy) e mitigazioni basate su livelli di maturità. Le PA devono mappare i controlli del NIST AI RMF su quattro funzioni: Govern, Map, Measure, Manage.
- **ISO 42001**: certificazione per i sistemi di gestione dell'IA, con focus su accountability e tracciabilità delle decisioni algoritmiche. Richiede l'identificazione di 39 controlli nell'Annex A, tra cui auditabilità dei modelli e monitoraggio dei *bias*.
- **MIT AI Risk Repository**: catalogazione dei rischi specifici per contesti pubblici (es. discriminazione in servizi sociali). Da utilizzare per scenari di stress-testing.

Checklist operativa:

- Condurre un gap analysis tra i controlli ISO 42001 e i processi esistenti, con priorità sulle lacune ad alto impatto.
- Adottare il Generative AI Profile del NIST per sistemi come chatbot, con test su 15 parametri di robustezza.
- Integrare nel registro IA un modulo per la tracciabilità dei dataset, includendo provenienza, licenze e potenziali *bias*.

1.3 Creazione di un AI Governance Board

La creazione di un **AI Governance Board** (AGB) rappresenta un passo cruciale per garantire che i sistemi di IA siano progettati e implementati rispettando i principi di trasparenza, equità e responsabilità. L'AGB dovrebbe includere:

- **Dirigenti della PA**: per assicurare che il progetto sia allineato agli obiettivi strategici dell'amministrazione.
- **Esperti legali**: per monitorare la conformità normativa, in particolare riguardo al GDPR e all'AI Act.
- **Rappresentanti etici**: provenienti da ONG, università e associazioni civiche, per supervisionare l'aderenza ai principi etici.
- **Stakeholder tecnici**: ingegneri e data scientist per fornire supporto operativo e garantire la robustezza tecnica del sistema.

Compiti principali dell'AI Governance Board

1. Definizione delle policy sull'uso dell'IA

- Stabilire linee guida per l'utilizzo responsabile dell'IA, come la necessità di rendere spiegabili tutte le decisioni automatizzate.
- Adottare criteri di supervisione umana per le decisioni ad alto impatto.

2. Monitoraggio degli impatti sociali ed economici

- Analizzare come le decisioni automatizzate influenzano gruppi vulnerabili e verificare che non si creino discriminazioni sistemiche.
- Valutare i benefici economici derivanti dall'automazione, come la riduzione del carico amministrativo.

3. Supervisione dei KPI

- Stabilire indicatori di performance come:
  - **Accuratezza delle decisioni**: Maggiore del 97%.
  - **Percentuale di decisioni contestate**: Inferiore al 5%.
  - **Tempo medio di risposta**: Inferiore a 24 ore.

1.4 Policy e linee guida operative

Le policy operative devono essere documentate e aggiornate regolarmente per garantire che il sistema di IA rispetti le normative e i principi etici. Le linee guida includono:

- **Trasparenza**: ogni decisione automatizzata deve essere accompagnata da una spiegazione comprensibile per l'utente, utilizzando strumenti di Explainable AI (XAI).
- **Equità**: i dataset utilizzati devono essere rappresentativi della popolazione servita per prevenire discriminazioni indirette.
- **Protezione dei dati**: devono essere implementate misure di sicurezza avanzate, come l'anonimizzazione e la crittografia dei dati personali.
- **Supervisione umana**: le decisioni che superano soglie di rischio predefinite devono essere esaminate manualmente da operatori esperti.

2. Valutazione del Rischio e Compliance Operativa

2.1 Modello di Valutazione Integrata (DPIA + FRIA)

L'AI Act richiede una **Fundamental Rights Impact Assessment (FRIA)** per i sistemi ad alto rischio, da condurre in parallelo alla DPIA del GDPR. Per la PA, questo si traduce in:

1. **Mappatura degli stakeholder**: cittadini, dipendenti, autorità di vigilanza. Utilizzare matrici di coinvolgimento basate sul livello di impatto (es. alto per sistemi di welfare).
2. **Analisi dei dati**: verifica della liceità del trattamento (base giuridica, minimizzazione dati, consenso esplicito per categorie speciali). Per dataset storici, applicare tecniche di *debiasing* come *reweighting* o *adversarial debiasing*.
3. **Valutazione del *bias***: utilizzo di tool come AI Fairness 360 (IBM) per rilevare discriminazioni in dataset storici. Soglia di disparità accettabile: <5% tra gruppi protetti.
4. **Test di robustezza**: simulazione di attacchi *adversarial* per verificare la resilienza dei modelli. Per sistemi NLP, testare con prompt injection e dati *out-of-distribution*.

Checklist operativa:

- Documentare la base giuridica per ogni trattamento (GDPR Art.6), specificando se si applicano eccezioni per 'interesse pubblico' (Art. 6(1)(e)).
- Implementare meccanismi di *anonymization* o *pseudonymization* per dataset sensibili, con valutazione del rischio di re-identificazione tramite k-anonymity (k≥3).
- Redigere report di conformità AI Act (Allegato III) per sistemi ad alto rischio, includendo piani di mitigazione per almeno tre scenari di *failure*.

2.2 Human-in-the-loop e spiegabilità

L'art. 22 GDPR e l'AI Act impongono il diritto a un intervento umano su decisioni automatizzate. Nella PA, ciò si applica a:

- **Processi amministrativi**: es. assegnazione di benefici sociali, con dashboard di monitoraggio in tempo reale che segnalano anomalie statistiche (es. deviazione >2σ dalla media storica).
- **Comunicazione con i cittadini**: chatbot devono chiarire quando non sono in grado di rispondere e attivare canali umani. Soglia di incertezza: confidence score <85%.

Checklist operativa:

- Integrare dashboard di monitoraggio con metriche di *fairness* (disparità di trattamento per gruppi demografici) aggiornate ogni 24 ore.
- Fornire spiegazioni 'a gradiente' (*layer-wise relevance propagation*) per decisioni critiche, con report generati in formato XAI (XML for AI Explanations).
- Addestrare dipendenti su protocolli di *override* dei sistemi IA, con simulazioni semestrali su casi reali (es. correzione manuale di assegnazioni errate di alloggi popolari).

3. Sicurezza Cibernetica e Resilienza

3.1 Protezione dei Sistemi IA

L'AI Act richiede che i sistemi ad alto rischio siano progettati con sicurezza by design. Per la PA:

- **Modello di threat intelligence**: collaborazione con CERT-AgID per identificare attacchi mirati (es. *data poisoning*). Implementare *honeypot* con dati sintetici per rilevare tentativi di manipolazione.
- **Crittografia omomorfica**: per elaborare dati sensibili senza esporli in chiaro. Utilizzare librerie verificabili (es. Microsoft SEAL) con audit di sicurezza semestrali.
- **Controlli di accesso RBAC (Role-Based)**: limitazione dei privilegi in base al principio del minimo privilegio. Ruoli predefiniti: Data Scientist (sola lettura modelli), Auditor (accesso completo), Operatore (solo inferenza).

Checklist operativa:

- Eseguire penetration test semestrali su API e modelli esposti, utilizzando framework come Adversarial Robustness Toolbox.
- Utilizzare hardware *trusted execution environment* (TEE) per modelli critici, con certificazione Common criteria EAL4+.
- Implementare SIEM (Security Information and Event Management) con regole specifiche per *anomaly detection* IA, come variazioni >30% nell'*accuracy* di modello.

3.2 Gestione degli incidenti

In caso di *data breach* o malfunzionamenti algoritmici, la PA deve attivare protocolli conformi al GDPR Art. 33-34 e all'AI Act Art. 17:

- **Notifica all'Autorità garante**: entro 72 ore dalla rilevazione, con descrizione del sistema IA coinvolto, impatto stimato su diritti fondamentali, e misure correttive immediate.
- **Comunicazione agli interessati**: se il *breach* comporta rischi per diritti e libertà. Utilizzare canali prioritari (SMS, e-mail certificata) con linguaggio chiaro (livello A2 QCER).
- **Post-mortem analitico**: identificazione delle root cause tecniche e organizzative. Per incidenti gravi (>10.000 persone coinvolte), coinvolgere esperti terzi indipendenti 4.

Checklist operativa:

- Mantenere un inventario aggiornato dei contatti delle autorità competenti (Garante privacy, AGID) in formato *machine-readable* (es. JSON-LD).
- Predisporre template precompilati per le notifiche, con campi obbligatori indicati dall'AI Act Annex V[2].
- Integrare nel SIEM un modulo di *forensic automation* per la raccolta automatica di log rilevanti (es. modifiche ai modelli nelle ultime 72 ore).

4. Formazione e Cambiamento culturale

La PA deve superare il digital divide interno attraverso:

- **Corsi certificati su AI Ethics**: in collaborazione con università (es. Università di Genova) e organismi come AgID. Durata minima: 40 ore, con esame finale su casi reali.
- **Simulazioni di caso**: esercitazioni su scenari reali (es. correzione di bias in algoritmi di assegnazione alloggi) utilizzando piattaforme sandbox con dati sintetici.
- **Toolkit per Dipartimenti**: linee guida su prompt engineering per sistemi generativi (es. ChatGPT per redazione atti), con whitelist di termini vietati (es. "cittadino onorevole").

Checklist operativa:

- Definire un piano formativo annuale con KPI di completamento (minimo 80% del personale entro Q4).
- Istituire comunità di pratica tra enti locali per condividere best practice.

5. Processo di Gestione dei rischi

5.1. Identificazione e Classificazione dei Rischi

La gestione dei rischi è un elemento cruciale nella progettazione e implementazione di sistemi di Intelligenza artificiale (IA) nelle Pubbliche Amministrazioni (PA). L'identificazione e la classificazione dei

2 Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024.

rischi mirano a prevenire problematiche etiche, tecniche e operative, garantendo che il sistema sia equo e conforme alle normative.

Rischi etici

- **Discriminazione indiretta**: la presenza di bias nei dati di addestramento può causare trattamenti ingiusti per gruppi specifici di utenti.
- **Mancanza di trasparenza**: decisioni non spiegabili possono ridurre la fiducia degli utenti e compromettere la responsabilità amministrativa.

Rischi tecnici

- **Attacchi informatici**: manipolazione intenzionale dei dati di input (data poisoning) per influenzare i risultati.
- **Errori sistematici**: correlazioni spurie nei dataset che portano a decisioni errate.

Rischi operativi

- **Scarsa supervisione**: mancanza di competenze tecniche tra il personale amministrativo per gestire le complessità del sistema.
- **Incapacità di adattamento**: difficoltà nel rispondere rapidamente a cambiamenti normativi o sociali.

5.2. Piano di Mitigazione

Un piano di mitigazione strutturato è essenziale per affrontare i rischi identificati, riducendo l'impatto negativo e aumentando la resilienza del sistema. Le principali attività includono:

Validazione dei dataset

- **Obiettivo**: garantire che i dati utilizzati siano rappresentativi e bilanciati.
- **Esempio pratico**: una PA ha riscontrato che un modello per distribuire fondi alle imprese favoriva le grandi aziende urbane. Per correggere questa disparità, è stato introdotto un filtro per normalizzare i dati in base a settore e dimensione aziendale, migliorando l'equità nelle assegnazioni.

Simulazioni e test di robustezza

- **Obiettivo**: verificare il comportamento del sistema in scenari critici.
- **Esempio pratico**: test di simulazione hanno evidenziato che il sistema era vulnerabile a manipolazioni dei dati di input. Sono stati quindi implementati controlli di integrità e meccanismi di rilevamento per prevenire tali manipolazioni.

Strumenti utilizzati

- **Fairlearn**: per identificare e mitigare i bias nei dataset.
- **What-If Tool**: per simulare scenari ipotetici e analizzare l'impatto delle decisioni del modello.

5.3. Monitoraggio continuo

Il monitoraggio continuo garantisce che il sistema mantenga performance ottimali e conformità alle normative, adattandosi alle esigenze emergenti.

Le linee guida AgID introducono un **Maturity Model** su cinque livelli (Initial, Managed, Defined, Measured, Optimizing), aiutando gli enti a autovalutarsi. Parametri chiave includono:

- **Allineamento Strategico**: integrazione dell'IA negli obiettivi istituzionali.
- **Gestione del Rischio**: frequenza di aggiornamento delle FRIA.
- **Impatto Sociale**: metriche di soddisfazione cittadina post-implementazione IA.

Le modalità tramite cui possono essere individuati i punti da migliorare possono essere:

1. Audit Trimestrali

- **Obiettivo**: valutare regolarmente l'equità, l'accuratezza e la trasparenza del sistema.
- **Esempio pratico**: un sistema per l'assegnazione di borse di studio viene sottoposto ad audit trimestrali per identificare variazioni nei criteri di selezione che potrebbero indicare la presenza di bias.

2. Raccolta di Feedback dai Cittadini

- **Obiettivo**: coinvolgere gli utenti finali per migliorare il sistema sulla base delle loro esperienze e segnalazioni.
- **Esempio pratico**: una PA ha implementato un'app dedicata che consente agli utenti di contestare decisioni errate o richiedere chiarimenti. Le segnalazioni raccolte sono state utilizzate per migliorare l'accuratezza del modello.

3. Revisione iterativa dei Modelli

- **Obiettivo**: aggiornare periodicamente i modelli per rispondere a cambiamenti nei contesti normativi e sociali.
- **Esempio pratico**: una PA aggiorna il modello utilizzato per l'assegnazione di contributi agricoli per includere le nuove normative europee sui sussidi verdi.

5.4. Processo di Audit

Un elemento cruciale per garantire la conformità e il miglioramento continuo dei sistemi di Intelligenza artificiale (IA) nella Pubblica amministrazione (PA) è rappresentato dal processo di audit. Questo capitolo descrive le diverse fasi dell'audit, che includono verifiche iniziali, periodiche e straordinarie, illustrando casi concreti e approcci metodologici per una corretta implementazione.

5.4.1. Audit Iniziale

Obiettivo

L'audit iniziale mira a garantire che il sistema di IA sia conforme ai requisiti normativi, tecnici ed etici prima della sua implementazione operativa.

Attività principali

1. Validazione della qualità dei dati

- **Descrizione**: i dataset utilizzati per addestrare il modello devono essere rappresentativi, completi e privi di bias significativi.
- **Caso pratico**: una PA che sviluppa un sistema per il calcolo delle tasse ha verificato che i dati storici includessero correttamente tutte le fasce di reddito e non presentassero anomalie che potessero influire sull'equità delle decisioni.

2. Test di robustezza del modello

- **Descrizione**: il modello deve essere sottoposto a test per valutare la sua capacità di gestire input inattesi o incompleti.
- **Caso pratico**: la PA ha eseguito test simulando errori nei dati di input, come valori fuori scala o incompleti, per garantire che il sistema rispondesse con risultati affidabili o generasse avvisi appropriati.

3. Valutazione dei rischi etici

- **Descrizione**: un'analisi dettagliata deve identificare potenziali impatti etici, come discriminazioni indirette o mancanza di trasparenza.
- **Caso pratico**: durante l'audit di un sistema di assegnazione di borse di studio, è stato verificato che il modello non penalizzasse inconsapevolmente studenti provenienti da scuole situate in aree svantaggiate.

5.4.2. Audit Periodico

Obiettivo

Gli audit periodici garantiscono che il sistema di IA continui a operare in modo conforme, identificando eventuali nuove problematiche emerse durante il suo utilizzo.

Attività Principali

1. Analisi Campionata delle Decisioni

- **Descrizione**: viene analizzato un campione rappresentativo delle decisioni del sistema per verificare l'equità e l'aderenza alle policy operative.
- **Caso pratico:** una PA che gestisce un sistema per l'assegnazione di contributi culturali controlla trimestralmente che le decisioni siano distribuite equamente tra enti pubblici e privati, evitando concentrazioni eccessive.

2. Aggiornamento del Registro dei Rischi

- **Descrizione**: le valutazioni periodiche dei rischi vengono integrate con nuove informazioni derivanti dall'uso reale del sistema.

- **Caso pratico:** durante un audit trimestrale, la PA ha rilevato che il sistema stava privilegiando imprese con maggiori risorse tecniche per completare correttamente la domanda online. Questo ha portato all'introduzione di una guida interattiva per supportare le imprese meno tecnologicamente preparate.

3. Monitoraggio dei KPI

- **Descrizione**: gli indicatori di performance vengono confrontati con le soglie target per garantire che il sistema operi entro i parametri stabiliti.
- **Caso pratico**: un sistema di gestione dei trasporti pubblici utilizza KPI come il tempo di risposta alle segnalazioni degli utenti e la precisione nelle previsioni di affluenza.

5.4.3. Audit Straordinario

Obiettivo

L'audit straordinario viene attivato in risposta a incidenti critici o segnalazioni significative, con l'obiettivo di identificare e risolvere le cause profonde dei problemi.

Attività Principali

1. Analisi Forense dei Log

- **Descrizione**: vengono analizzati i registri delle attività del sistema per tracciare le decisioni problematiche e identificare eventuali anomalie.
- **Caso pratico:** in seguito a una segnalazione di errori nell'assegnazione delle borse di studio, la PA ha analizzato i log e scoperto che un parametro mal configurato aveva penalizzato alcune richieste legittime.

2. Revisione dei Processi Operativi

- **Descrizione**: l'intero processo decisionale del sistema viene riesaminato per individuare eventuali lacune nelle policy o nelle procedure.
- **Caso pratico:** dopo un audit straordinario, una PA ha introdotto controlli aggiuntivi per garantire che le modifiche ai parametri del modello fossero approvate da un team multidisciplinare.

3. Proposta di Miglioramenti

- **Descrizione**: sulla base dei risultati dell'audit, vengono sviluppate raccomandazioni per prevenire il ripetersi degli incidenti e migliorare la resilienza del sistema.
- **Caso pratico:** una PA ha implementato una dashboard di monitoraggio in tempo reale per rilevare automaticamente eventuali discrepanze nelle decisioni del sistema.

Bibliografia

Bannister F., Connolly R. (2014), ICT, public values and transformative government. A framework and programme for research, *Government Information Quarterly*, 31, n.1, pp.119-128

Criado J.I., Gil-Garcia J.R. (2019), Creating public value through smart technologies and strategies: From digital services to artificial intelligence and beyond, *International Journal of Public Sector Management*, 32, n.5, pp.438-450

Cugurullo F. (2021), Urban artificial intelligence: From automation to autonomy in the smart city, *Frontiers in Sustainable Cities*, n.3, Article 38

de Bruijn H., Janssen M., Warnier M. (2020), Accountability challenges in the transnational algorithmic governance: The case of public sector algorithms, in Ebers M., Cantero Gamito M. (eds), *Algorithmic Governance and Governance of Algorithms: Legal and Ethical Challenges,* Berlin, Springer, pp.137-156

Floridi L., Cowls J. (2019), A unified framework of five principles for AI in society, *Harvard Data Science Review*, 1, n.1 <DOI:10.1162/99608f92.8cd550d1>

Floridi L., Cowls J., King T. C., Taddeo M. (2020), How to design AI for social good: Seven essential factors, *Science and Engineering Ethics*, 26, n.3, pp.1771-1796

Janowski T., Estevez E., Ojo A. (2022), Digital government for sustainable development, *Government Information Quarterly*, 39, n.2, 101667

Janssen M., Brous P., Estevez E., Barbosa L.S., Janowski T. (2020), Data governance: Organizing data for trustworthy Artificial Intelligence, *Government Information Quarterly*, 37, n.3, 101493

Janssen M., van der Voort H. (2016), Adaptive governance: Towards a stable, accountable and responsive government, *Government Information Quarterly*, 37, n.1, 101411

Kaminski M.E., Malgieri G. (2021), Algorithmic impact assessments under the GDPR: Producing multi-layered explanations, *International Data Privacy Law*, 11, n.2, pp.125-144

Lewis D., Mooney J., Brady M., Sánchez-Ortiz V., Litchfield I. (2022), The adoption of artificial intelligence in public services: A systematic literature review, *Public Administration Review*, 82, n.6, pp.1049-1064

Lodge M., Mennicken A. (2019), The importance of regulation of and by algorithm, in Yeung M., Lodge K. (eds.), *Algorithmic Regulation*, Oxford, Oxford University Press, pp.1-12

Mergel I., Edelmann N., Haug N. (2019), Defining digital transformation: Results from expert interviews, *Government Information Quarterly*, 36, n.4, 101385

Misuraca G., van Noordt C., Boukli A. (2020), The use of AI in public services: Results from a preliminary mapping across the EU, in *Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance (ICEGOV 2020)*, New York, ACM, pp.90-99

Moher D., Liberati A., Tetzlaff J., Altman D.G. (2009), Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement, *British Medical Journal*, n.339, b2535

Moore M.H. (1995), *Creating public value: Strategic management in government*, Harvard, Harvard University Press

ÒReilly T. (2011), Government as a platform. Innovations: Technology, *Governance, Globalization*, 6, n.1, pp.13-40

Radford A., Wu J., Child R., Luan D., Amodei D., Sutskever I. (2019), Language models are unsupervised multitask learners, *OpenAI Blog*, 1, n.8, p.9

Sun T.Q., Medaglia R. (2019), Mapping the challenges of artificial intelligence in the public sector: Evidence from public healthcare, *Government Information Quarterly*, 36, n.2, pp.368-383

Twizeyimana J.D., Andersson A. (2019), The public value of E-Government. A literature review, Government Information Quarterly, 36, n.2, pp.167-178

van der Voort H.G., Klievink A.J., Arnaboldi M., Meijer A.J. (2019), Rationality and politics of algorithms, Will the promise of big data survive the dynamics of public decision making?, *Government Information Quarterly*, 36, n.1, pp.27-38

Veale M., Borgesius F.Z. (2021), Demystifying the Draft EU Artificial Intelligence Act, *Computer Law Review International*, 22, n.4, pp.97-112

Venkatesh V., Thong J.Y., Xu X. (2016), Unified theory of acceptance and use of technology: A synthesis and the road ahead, *Journal of the Association for Information Systems*, 17, n.5, pp.328-376

Wirtz B.W., Müller W.M. (2023), Artificial Intelligence and public sector management: A systematic literature review, *International Journal of Public Administration*, 46, n.1, pp.34-51

Wirtz B. W., Weyerer J. C., Geyer C. (2022), Artificial intelligence in the public sector: Foundations, applications, consequences, and challenges, *International Journal of Public Administration*, 45, n.1, pp.80-89

Yin R.K. (2018), *Case study research and applications: Design and methods*, Los Angeles, Sage Publications

Gaetano Riccio
gaetano.riccio84@gmail.com
È Docente a contratto presso l'Università degli Studi di Napoli Federico II, corso di Sistemi di elaborazione delle informazioni, e presso l'Università degli Studi della Campania Luigi Vanvitelli, corso Sistemi informativi per la PA. È Dirigente dell'Area Informatica della Giunta della Regione Campania, Direttore del Dipartimento Intelligenza artificiale di supporto alle Pubbliche amministrazioni dell'Ente nazionale Intelligenza artificiale (ENIA) e Componente esterno del Comitato d'indirizzo dei Corsi di studio dell'Area economica del Dipartimento di Economia, management e metodi quantitativi.

Banche, finanza e assicurazioni: sfide per i lavoratori e per le politiche pubbliche nell'era dell'Intelligenza artificiale

Andrea Battistoni
INAPP

Valentina Ferri
INAPP

Lo studio analizza l'impatto dell'IA sull'occupazione nel settore bancario, finanziario e assicurativo in Italia, combinando indici di esposizione e dati delle Forze di lavoro e delle Comunicazioni Obbligatorie. Dalle analisi descrittive ed econometriche emergono segnali di riduzione occupazionale legati alla maggiore esposizione tecnologica. L'analisi suggerisce l'urgenza di politiche di *upskilling* e *reskilling* per accompagnare la trasformazione del settore e tutelare i lavoratori più a rischio.

The study analyzes the impact of Artificial intelligence on employment in the banking, financial, and insurance sectors in Italy, combining exposure indices with data from the Labour Force Survey and Mandatory Communications. Descriptive and econometric analyses reveal signs of employment reduction linked to higher technological exposure. The findings highlight the urgency of upskilling and reskilling policies to support the sector's transformation and protect the most vulnerable workers.

DOI: 10.53223/Sinappsi_2025-02-12

Citazione
Battistoni A., Ferri V. (2025), Banche, finanza e assicurazioni: sfide per i lavoratori e per le politiche pubbliche nell'era dell'Intelligenza artificiale, *Sinappsi*, XV, n.2, pp.171-191

Parole chiave
Banca finanza e assicurazioni
Intelligenza artificiale
Mercato del lavoro

Keywords
Banking, financial, and insurance sectors
Artificial intelligence
Labour Market

Introduzione

Secondo un recente studio sui settori che mediamente saranno impattati maggiormente dall'Intelligenza artificiale (IA) in Italia, emerge che nell'ambito bancario, finanziario e assicurativo si trovano lavoratori maggiormente a rischio di sostituzione (Battistoni e Ferri 2024). È opportuno, infatti, evidenziare che le tecnologie hanno già iniziato da anni un'ascesa incalzante in tale settore, influendo in modo incisivo sulla struttura occupazionale. Questa ascesa è molto visibile proprio a seguito dell'introduzione di alcune applicazioni quali ad esempio l'introduzione degli ATM e, successivamente, l'home banking. L'introduzione di queste innovazioni ha comportato un ridimensionamento dei luoghi fisici in cui effettuare le operazioni bancarie. Un elemento che da alcuni anni lascia presagire la progressiva riduzione della necessità di lavoratori è l'introduzione di chatbot e assistenti virtuali, sostitutivi in molti casi di individui che realizzano determinate operazioni, in grado di offrire risposte rapide 24 ore su 24.

Anche nel settore assicurativo, l'avvento delle nuove tecniche garantisce performance più elevate. Si individua, ad esempio, una migliore gestione dell'analisi del rischio e la tecnologia è utilizzata per la creazione di modelli predittivi che possano stimare il comportamento eventualmente insolvente dei

clienti. Il settore della finanza risulta anch'esso fortemente influenzato dalle tante applicazioni che possono essere utilizzate sostituendosi nelle attività al contributo normalmente dato dal lavoratore al perseguimento dei risultati. Un elemento che distingue la diffusione dell'IA rispetto a molte altre tecnologie innovative introdotte in passato riguarda il diverso impatto sul mercato del lavoro. Con la diffusione dell'IA avviene il passaggio da un'epoca in cui le mansioni più routinarie potevano essere automatizzate più facilmente, a un'epoca in cui è possibile effettuare attività decisamente più complesse come stime, creazione di modelli predittivi e fornire risposte personalizzate ai clienti. Questo cambiamento sposta le preoccupazioni del dibattito scientifico dalla sostituzione delle mansioni più routinarie alla possibile sostituzione di mansioni di livello più elevato.

Il contributo si apre con una rassegna della letteratura volta ad approfondire il processo di transizione dall'automazione tradizionale all'IA e ad esaminare le implicazioni di questo cambiamento sul mercato del lavoro. Successivamente si passano in rassegna tutte le innovazioni che hanno riguardato segnatamente il settore bancario, assicurativo e finanziario negli anni più recenti. Il paragrafo riguardante i dati e la metodologia permetterà di illustrare le fonti su cui si baseranno le analisi descrittive. Inoltre, nella stessa sezione verranno spiegati i metodi con cui sono stati costruiti gli Indici di esposizione all'IA delle professioni (AIOE), nonché l'Indice dell'esposizione all'IA corretto (CAIOE). Si presenta poi una panoramica generale su come l'IA viene integrata nelle imprese e nei processi che caratterizzano queste ultime e, sulla base di questi Indici, vengono condotte delle analisi descrittive nonché delle analisi statistiche ed econometriche con le quali si analizzano aspetti occupazionali relativi al settore oggetto di studio.

1. Dall'automazione all'Intelligenza artificiale: letteratura sull'impatto nel mercato del lavoro

Nell'analizzare il rapporto tra ascesa di nuove tecnologie e occupazione, si tende a includere gli studi legati all'automazione nonché alla routinarietà delle mansioni, che hanno cambiato il volto della produttività e hanno modificato la struttura del mercato del lavoro. Questo approccio comparativo permette di confrontare le precedenti tendenze relative ai fenomeni che hanno seguito l'ascesa dell'automazione con quelle che potrebbero emergere con l'introduzione dell'IA. Divisione del lavoro e automazione delle macchine sono state negli anni grandi driver che hanno comportato l'aumento della produttività. Ernst *et al.* (2019) hanno tracciato un quadro che indica come l'ascesa delle tecnologie ha modificato il mercato del lavoro negli anni passati, identificando come i fenomeni si siano collocati precisamente nei vari periodi storici. Gli autori hanno evidenziato come l'automazione dagli anni Cinquanta agli anni Settanta abbia spostato la manodopera dall'agricoltura alla manifattura e alle costruzioni, aumentando la domanda di lavoratori qualificati e ampliando il divario con quelli meno qualificati. Precedentemente all'introduzione dei sistemi di IA, un punto nodale degli studi nella letteratura degli economisti del lavoro riguardava, pertanto, l'esposizione delle professioni alla routinarietà. L'automazione è stato un argomento trattato in relazione al potenziale rischio di perdita di posti di lavoro da Frey e Osborne (2017) che hanno rivelato che il 47% di 702 occupazioni negli Stati Uniti erano ad alto rischio di automazione, in particolare nei settori dei trasporti e della produzione non qualificata.

In un'ottica di compensazione delle nuove tecnologie, che porterebbero da una parte alla distruzione di posti di lavoro e dall'altra ad un'eventuale creazione di occupazione, è possibile trovare in letteratura studi che evidenziano anzitutto un calo degli occupati in tutte le principali economie avanzate, dove i compiti di routine e ripetitivi svolti dall'essere umano sono stati sostituiti da robot e macchine di automazione industriale. D'altro canto, attraverso la progettazione e l'implementazione dei robot, sembrerebbe essere nata una nuova economia che ha potuto incidere nel lungo periodo a tal punto da far emergere effetti positivi sull'occupazione (Acemoglu e Restrepo 2017; Bessen 2017; Chiacchio *et al.* 2018; De Backer e DeStefano 2021; Graetz e Michaels 2015; Ernst *et al.* 2019). L'analisi relativa ai Paesi in via di sviluppo, invece, mostrerebbe che l'introduzione dei robot incide negativamente e significativamente sull'occupazione (Carbonero *et al.* 2018).

Probabilmente gli effetti negativi e significativi riscontrati potrebbero essere dovuti alle tecnologie adottate in tempi più lunghi da parte dei Paesi in via di sviluppo, i quali non hanno ancora esaurito le tre fasi che la letteratura individua (Acemoglu

e Restrepo 2017; Chiacchio *et al.* 2018; Vivarelli 2014). Invero, i passaggi che molti studi tracciano a seguito dell'introduzione di nuove tecnologie sono i seguenti: la disoccupazione inizialmente aumenta con l'automazione prima di diminuire nuovamente quando i prezzi e la produttività si adatteranno in una fase successiva. Le tre fasi sono così caratterizzate: in prima battuta si ha una sostituzione diretta dei lavori e dei compiti attualmente svolti dai lavoratori (effetto di sostituzione); successivamente, si evidenzierebbe un aumento complementare dei lavori e dei compiti necessari per utilizzare, gestire e supervisionare le nuove macchine (effetto di complementarità delle competenze); e in terzo luogo, si osserva un effetto di domanda determinato sia dai prezzi più bassi, sia da un aumento generale del reddito disponibile nell'economia grazie alla maggiore produttività (effetto di produttività). Benché la letteratura individui una prima fase di cosiddetta sostituzione dei lavori, ciò che in prima battuta si sostituisce sono alcuni *tasks* che il lavoratore svolge, e questo determinerebbe una probabilità di sostituzione del lavoratore non nella totalità dei compiti svolti, ma in una parte di essi, almeno nel breve periodo. Va altresì evidenziato che in tale dinamica si potrebbe individuare una sorta di progressività della sostituzione. L'avanzare delle tecnologie e il suo effetto sull'occupazione sarebbero legati alla possibilità di soppiantare il lavoratore in alcune piuttosto che in tutte le aree di attività che svolge nella sua giornata tipo. Sarebbe quindi utile, in ogni studio, includere la variabile del tempo che non può considerarsi neutrale nell'impatto sul mercato del lavoro e costituisce un ovvio limite degli studi che considerano le tecnologie ad oggi senza conoscerne in alcun modo l'evolversi.

Proprio in questa direzione va il lavoro di Huang e Rust (2018), i quali sostengono che l'IA nel trasformare i servizi costituisce una minaccia per i posti di lavoro. La teoria della sostituzione dei lavori da parte dell'IA individua quattro tipi di intelligenza necessari per lo svolgimento dei *tasks* legati ai servizi. Si parte dall'intelligenza meccanica, seguita da quella analitica, poi da quella intuitiva e successivamente da quella empatica. Secondo lo studio, attraverso la suddetta segmentazione delle intelligenze, è possibile comprendere il modo in cui le aziende dovrebbero decidere tra esseri umani e macchine per svolgere determinati *tasks*. La teoria afferma che, inizialmente, i lavoratori vengono sostituiti per i compiti che richiedono forme di 'intelligenza inferiore', ovvero più facilmente replicabili dall'IA. In questo modo, la sostituzione del lavoro umano potrebbe avvenire in modo graduale.

Un'importante implicazione di questa teoria è che le competenze analitiche tenderanno a perdere importanza nel tempo, poiché l'IA assumerà una quota sempre maggiore di compiti analitici, accrescendo invece il valore delle competenze 'soft', intuitive ed empatiche possedute dai dipendenti dei servizi.

Infine, l'IA diventerà capace di svolgere anche compiti intuitivi ed empatici e si arriverebbe, così, alla vera e propria minaccia per l'occupazione umana (Huang e Rust 2018). Secondo Ernst *et al.* (2019) i lavori sono costituiti da una serie di *tasks*, nel momento in cui alcuni di questi *tasks* risulteranno automatizzati, i profili di lavoro potranno cambiare. L'aggiunta di nuovi *tasks* e la modifica di quelli esistenti, piuttosto che eliminare completamente un lavoro, ne cambierebbe tuttavia il volto. Un esempio evidente è la descrizione del ruolo di un assistente amministrativo. Tale lavoratore nel tempo può dimostrare come si possa continuare a svolgere determinati *tasks* che non sono stati (ancora) automatizzati, insieme a nuovi compiti che non esistevano prima o che eventualmente erano eseguiti da un diverso gruppo di lavoratori. Pertanto, secondo tale ricerca, la scomparsa dei lavori dipende dal fatto che sia ancora utile raggruppare determinati compiti in profili di lavoro specifici e assumere lavoratori specificamente per questi (nuovi) lavori. Da quanto fin qui evidenziato, sembrerebbe emergere la necessità di realizzare gli studi relativi all'IA tenendo conto dei *tasks* che caratterizzano il lavoro di ogni individuo, inoltre, sembrerebbe evidenziarsi la necessità di uscire dai confini legati alla tipica classificazione professionale che potrebbe non contenere al suo interno tutti gli elementi utili alla definizione di come cambiano le attività svolte dal lavoratore.

L'Inapp implementa uno strumento molto prezioso a cui è legato tutto il patrimonio relativo alla formazione e alla possibilità di individuare, validare e certificare le competenze: l'Atlante del Lavoro e delle qualificazioni (Mazzarella *et al.* 2017). In questo caso, è stato superato l'approccio che corrisponde a una classificazione statistica ma è possibile segmentare il lavoro in aree di attività, unendo caratteristiche della professione al settore ATECO e

individuando le singole attività e i risultati attesi. Ciò permette di fare molte analisi e legare le *vacancies* ad una serie di altri dati (Camassa *et al.* 2024) e potrebbe pertanto essere utile anche nell'analisi che di seguito si propone[1]. È opportuno evidenziare tuttavia che l'approccio relativo alle professioni è stato fortemente utilizzato in letteratura e, certamente, anche grazie alla possibilità di legarsi a moltissime banche dati può offrire validissime indicazioni in termini di sostituzione potenziale del lavoratore. Nella letteratura americana prima, e in quella europea successivamente, sono stati molti i lavori legati alle professioni e alla potenziale esposizione dei lavoratori. Attraverso valori che misurano le abilità che l'IA riuscirebbe a sostituire, è stato possibile costruire degli indicatori compositi con dei gradienti che indicherebbero quanto potenzialmente un lavoratore potrebbe essere esposto a sostituzione a causa di questa innovazione tecnologica.

Una delle misure più utilizzate in letteratura per studiare l'esposizione dei lavoratori all'IA si basa sul lavoro di Felten *et al.* (2021). Felten utilizza un questionario somministrato a 2.000 *Gig Workers* tramite il servizio web Mechanical Turk (mTurk) di Amazon. Gli intervistati rispondono su quanto ciascuna delle 52 abilità di O*NET potrebbe essere utile per l'utilizzo di 10 applicazioni tipiche dell'IA: giochi di strategia astratti, videogiochi in tempo reale, riconoscimento di immagini, risposte visive a domande, generazione di immagini, comprensione della lettura, modellazione linguistica, traduzione, riconoscimento vocale e riconoscimento di tracce strumentali[2]. Nel loro lavoro, Felten *et al.* (2021) individuano un punteggio di esposizione all'IA per ogni professione, specificando le tipologie di lavoratori più o meno esposti, senza fornire informazioni su un eventuale effetto complementare o sostitutivo dell'IA rispetto alla professione in questione (AIOE). Tali dati vengono poi aggregati e pertanto si ottiene un livello di esposizione dei settori (AIIE) e geografici (AIGE).

Pizzinelli *et al.* (2023) analizzano l'impatto dell'IA utilizzando l'Indice AIOE, concentrandosi maggiormente sui concetti di complementarità e sostituzione. Lo studio esplora in particolare la differenza tra complementarità, che aiuta l'individuo a svolgere varie attività, e sostituzione, che potrebbe portare a una diminuzione dei livelli occupazionali.

Il contributo propone un'estensione dell'Indice di Felten che incorpora la complementarità, portando alla creazione di un Indice corretto (C-AIOE). In questo Indice, l'esposizione alle occupazioni è corretta attraverso la loro potenziale complementarità con l'IA, rappresentata dal fattore theta: quanto più basso risulterebbe theta, tanto minore è l'esposizione.

Gli autori trovano che molti gruppi occupazionali altamente skillati con alta esposizione all'IA come 'manager' e 'professionals' hanno la potenziale complementarità più alta e l'Indice C-AIOE più basso.

Guarascio *et al.* (2023) analizzano le dinamiche del lavoro tra il 2011 e il 2018, dopo aver controllato per molti fattori relativi a domanda e offerta: gli autori trovano che, in media, l'esposizione all'IA ha un impatto positivo sull'occupazione regionale europea.

Lo studio di Ferri *et al.* (2024) colma un gap della letteratura poiché in esso si applicano gli indici di esposizione all'IA, comunemente utilizzati nella letteratura scientifica, per la prima volta al contesto italiano. Inoltre, si utilizzano classificazioni nazionali (ICP), superando l'approccio O*NET. Dalla stima dell'AIOE, emerge l'esposizione di varie professioni all'IA utilizzando il livello di abilità (AIOE). I ruoli con alta responsabilità, infatti, occupano una posizione tra i primi ranghi, suggerendo la necessità di rivedere l'indice per comprendere i potenziali benefici o effetti di sostituzione dell'IA sui lavoratori. In tal senso, lo studio introduce un Indice corretto (C-AIOE) sulla base di Pizzinelli *et al.* 2023 che considerano sei elementi per realizzare l'Indice theta: comunicazione, responsabilità, condizioni fisiche, criticità, routinizzazione e competenze. Un Indice theta più alto indica una maggiore complementarità con l'IA, mentre un indice più basso suggerisce un rischio maggiore di sostituzione. L'analisi rivela che il 23% dei lavoratori in Italia è a rischio significativo di essere sostituito dall'IA, in particolare coloro che ricoprono ruoli routinari con minori responsabilità e necessità di comunicazione, come impiegati generici e cassieri che sarebbero fortemente esposti al rischio di sostituzione.

1 Si veda: https://www.lavoro.gov.it/pagine/osservatorio-sulladozione-di-sistemi-di-intelligenza-artificiale-nel-mondo-del-lavoro.
2 Le applicazioni sono individuate dalla Electronic Frontier Foundation, fondata nel 1990.

D'altra parte, il 26,4% dei lavoratori è posizionato in modo tale da beneficiare dell'IA, indicando che, mentre alcuni lavori possono essere a rischio, altri potrebbero rivelare una maggiore produttività e supporto dalle tecnologie IA. Nell'occasional paper della Banca d'Italia (Dalla Zuanna *et al.* 2024) viene valutato l'impatto dell'Intelligenza artificiale sul mercato del lavoro italiano. Le tecnologie IA sono associate maggiormente alle occupazioni che richiedono competenze cognitive, in contrasto con le tecnologie di automazione precedenti che colpivano in maggior misura i lavori manuali. Questo cambiamento indica che i lavoratori impiegatizi potrebbero affrontare maggiori rischi dall'IA. Nel lavoro di ricerca si evidenzia che le occupazioni più esposte all'IA sono principalmente nel settore dei servizi che impiega un numero rilevante di lavoratori altamente qualificati. Questo settore è destinato a subire impatti più sostanziali dall'IA rispetto ai settori industriali e agricoli.

L'analisi si sofferma su una bassa mobilità lavorativa tra diversi tipi di occupazione, con i lavoratori in ruoli altamente esposti che spesso si spostano verso posizioni meno esposte, tipicamente con premi salariali inferiori. Anche in questo caso, tra i limiti del lavoro si evidenzia che le attuali misure di esposizione occupazionale si basano su definizioni statiche e, man mano che l'IA evolve, la natura dei lavori potrebbe cambiare, modificando il potenziale rischio di sostituzione o livello di complementarità in futuro.

Dagli studi finora evidenziati, che non riescono a separare la routinizzazione di molte attività svolte dai lavoratori dall'effetto complessivo dell'IA, emerge che l'IA può essere considerata un avanzamento tecnologico che, imitando il cervello umano, integra una serie di operazioni ripetitive svolte dagli individui. I compiti routinari, che sono evidentemente un sottoinsieme delle operazioni che l'IA può svolgere, possono essere quindi completamente sostituiti. L'insieme di tali compiti costituisce la base con cui l'IA realizza anche attività più complesse, avendo inglobato ulteriori competenze che permettono di analizzare dati e prendere decisioni.

D'altra parte, dalla review della letteratura emerge che le preoccupazioni riguardo all'ascesa delle nuove tecnologie sono una questione ricorrente in termini di impatto sui lavoratori. Sebbene nel breve periodo alcune considerazioni tendano a enfatizzare maggiormente gli effetti negativi, col passare degli anni e la sistematizzazione di alcuni processi legati a tali tecnologie, si possono spesso osservare effetti neutrali e positivi. Questo, nel caso dell'IA, potrebbe anche essere influenzato dalla diminuzione della forza lavoro dovuta a fattori demografici, che vedono ridursi negli anni la popolazione in età lavorativa. In ultima istanza sembrerebbe riscontrarsi una necessità di migliorare le competenze tecniche per adattarsi ai cambiamenti dell'IA. Il lavoro umano, infatti, si sposterà verso ruoli che richiedono competenze sociali e interpersonali.

In questo lavoro, applichiamo gli Indici sviluppati da Ferri *et al.* (2024) all'occupazione nel settore dei servizi bancari, finanziari e assicurativi, che sembra essere tra i più esposti all'ascesa dell'IA. Questo studio contribuisce alla letteratura proponendo una lettura dell'occupazione sulle Comunicazioni Obbligatorie (dati SISCO-MLPS), non solo per settore, ma anche per settore economico-professionale dell'Atlante del Lavoro passando per il tramite delle professioni. In questo modo, possiamo identificare nel primo caso gli occupati che svolgono qualsiasi professione nelle imprese di quel settore, mentre nel secondo caso osserviamo gli occupati che svolgono esclusivamente professioni coerenti con il settore stesso.

Inoltre, si osserva con attenzione il rapporto tra gli Indici finora sviluppati e le Comunicazioni Obbligatorie inerenti al settore delle banche, della finanza e delle assicurazioni, che non è stato ancora esplorato in letteratura. Questo studio sperimentale combina infine questi aspetti descrittivi innovativi con una lettura delle variazioni percentuali delle attivazioni dei contratti negli ultimi anni e un'analisi econometrica che mette in relazione tali variazioni percentuali con eventuali indici di esposizione all'IA. Inoltre, per implementare politiche pubbliche che possano mitigare i possibili effetti negativi dell'IA sul mercato del lavoro, è utile fornire una panoramica descrittiva dei trend occupazionali osservati negli anni, nonché delle transizioni lavorative tra diverse professioni nel settore. Le transizioni tra professioni con simile esposizione all'IA indicano una minore versatilità e un rischio maggiore per i lavoratori di non riuscire a collocarsi in profili professionali meno esposti all'IA, imponendo così una riflessione sull'*upskilling* e il *reskilling*.

I cambiamenti del settore bancario, assicurativo e finanziario

Negli ultimi anni, numerosi studiosi hanno esplorato gli avanzamenti tecnologici dei servizi, tra i lavori più recenti, alcuni mettono in luce l'importanza dell'Intelligenza artificiale nei servizi bancari. Ad esempio, Priya e Sharma (2023) analizzano l'efficacia delle applicazioni bancarie e dei meccanismi finanziari basati sull'IA. Rahman *et al.* (2021) sottolineano il ruolo cruciale dei chatbot assistenti IA. Singh (2019), invece esamina la qualità del servizio elettronico nel banking su Internet e la sua relazione con la soddisfazione del cliente. Kaneria (2022) discute come l'IA possa allineare i servizi finanziari attraverso un unico canale centrato sull'IA, e Paul *et al.* (2021) esplorano il ruolo dell'IA nei servizi bancari personalizzati.

L'IA, nei suddetti settori, migliora l'efficienza e i processi decisionali. Attraverso i più recenti sistemi tecnologici si automatizzano compiti ripetitivi, permettendo ai professionisti di concentrarsi su attività più strategiche; inoltre migliorano e vengono agevolati sia i processi di interazione con il cliente, nonché le previsioni sui dati (Huang e Rust, 2018). La letteratura identifica nel settore bancario e finanziario diversi aspetti principali che stanno trasformando radicalmente lo scenario lavorativo e le applicazioni che stanno maggiormente cambiando il panorama delle banche, delle assicurazioni e dei servizi finanziari in generale. La rapida diffusione di strumenti di IA generativa come i chatbot, come ChatGPT, creati dall'integrazione del *deep learning* e dei modelli linguistici basati sull'architettura Generative Pre-training Transformer (GPT), è destinata a cambiare il panorama degli agenti conversazionali (Thorp 2023). Questo sviluppo tecnologico richiede nuove conoscenze specifiche per comprendere la trasformazione delle app di IA nei servizi. I progressi nella robotica e nell'IA, insieme ai Big Data, stanno permettendo al settore bancario di introdurre app finanziarie di IA che offrono esperienze di servizio semi-personalizzate. Molte banche stanno adottando queste app per assistere i clienti nelle loro operazioni finanziarie, favorendo il passaggio dalle transazioni in contante a quelle digitali e migliorando il valore percepito dell'esperienza (Skandali *et al.* 2023). I chatbot permettono di migliorare il servizio clienti basandosi su grandi volumi di dati. L'introduzione di tali strumenti consente di ridurre i costi e gestire più clienti contemporaneamente, offrendo risposte in qualsiasi orario alle domande frequenti e monitorando sia le domande che le risposte dei clienti (Becerra-Vicario *et al.* 2024).

Tra gli strumenti più diffusi figurano i robo-advisor, utili nella gestione di portafogli di fondi di investimento e piani pensionistici, sebbene in alcuni casi, come nella prevenzione delle frodi, l'intervento umano resti indispensabile. Anche i bancomat hanno beneficiato dell'integrazione dell'IA, che ne ha potenziato la sicurezza e migliorato l'esperienza utente grazie a tecnologie come il riconoscimento facciale, la manutenzione predittiva e la previsione della domanda di contante. Tra le applicazioni più rilevanti si segnalano gli avvisi antifrode e le simulazioni personalizzate di prodotti finanziari.

Il mobile banking ha conosciuto una rapida crescita grazie ai dispositivi intelligenti che aiutano i consumatori a pianificare le finanze e a effettuare transazioni in modo efficiente. L'IA, in questo contesto, semplifica la gestione finanziaria personale (PFM) offrendo analisi personalizzate basate sulle abitudini di spesa e sugli obiettivi dell'utente. Inoltre, lo sviluppo di app che aggregano diversi prodotti finanziari consente una gestione semi-personalizzata del portafoglio tramite robo-advisor, riducendo progressivamente il bisogno dell'intermediazione umana in queste attività.

Nell'ambito delle applicazioni riconosciute da Becerra-Vicario *et al.* (2024) come i maggiori investimenti nei servizi finanziari e assicurativi, si annovera la possibilità di gestire i reclami valutando i modelli di comportamento degli utenti, riducendo i tempi di elaborazione e i costi. L'IA accelera inoltre i processi di sottoscrizione e liquidazione dei sinistri per le compagnie assicurative, raccogliendo dati estesi sui clienti per migliorare il servizio e lo sviluppo dei prodotti. Anche sugli aspetti relativi alla dinamica dei prezzi e alle valutazioni del rischio, l'IA offre un importante contributo. L'analisi predittiva permette di migliorare la strategia aziendale e l'ottimizzazione delle risorse, identificando modelli e prevedendo accuratamente il comportamento dei consumatori. Tecnologie come la blockchain e l'IA stanno ridefinendo il settore finanziario, offrendo soluzioni innovative per migliorare l'efficienza, la sicurezza e la gestione dei rischi. La blockchain affronta problematiche legate ai dati e alla prevenzione delle frodi (Becerra-Vicario *et al.* 2024).

Nel trading finanziario, l'IA analizza i dati storici per individuare tendenze e anomalie di mercato, offrendo soluzioni su misura in base ai profili di rischio degli investitori. Nella gestione patrimoniale, gli algoritmi intelligenti rendono i servizi di consulenza accessibili anche a clienti con piccoli patrimoni, automatizzando la gestione quotidiana delle finanze (Becerra-Vicario *et al.* 2024). Nel saggio di Kour e Kour (2024) viene fornito un esame completo di come il *machine learning* stia rimodellando l'industria finanziaria. In particolare, le pratiche tradizionali di analisi dei dati, decision-making e gestione del rischio in finanza sarebbero fortemente impattate. Questa trasformazione è poi evidente in varie applicazioni come la previsione del mercato azionario, la valutazione del rischio di credito, il rilevamento delle frodi, il trading algoritmico e l'ottimizzazione del portafoglio, che sfruttano grandi volumi di dati finanziari. Lo studio evidenzia diverse applicazioni di successo degli algoritmi di *machine learning* in finanza, tra cui la previsione del mercato azionario: attraverso l'utilizzo dei dati storici è possibile prevedere i prezzi futuri delle azioni. Successivamente tra le più importanti innovazioni viene citata la valutazione del rischio di credito, attraverso cui è possibile misurare l'affidabilità creditizia dei mutuatari utilizzando modelli predittivi.

Anche secondo lo studio di Farazi (2024) emerge che l'Intelligenza artificiale, in particolare il *machine learning* (ML) e le reti neurali artificiali (ANNs), migliora significativamente la gestione del rischio finanziario, permettendo analisi più accurate e individuando schemi complessi. A tal proposito nello studio si mettono in evidenza le limitazioni dei modelli tradizionali di gestione del rischio, come il VaR e lo stress testing, che risultano statici e dipendenti dai dati storici. Complessivamente, l'integrazione dell'IA nei framework di gestione del rischio può migliorare la previsione dei rischi e la stabilità finanziaria.

Dopo l'ascesa del banking online si sono evidenziate altre problematiche sul riciclaggio di denaro nonché frodi informatiche, e l'utilizzo dell'IA permette di rilevare più facilmente le attività illecite. Nell'ambito del credit scoring, l'IA consente la valutazione del rischio di default (Kour e Kour 2024). La valutazione della solvibilità dei mutuatari utilizzando modelli predittivi diventa un'operazione meno complessa con l'utilizzo del *machine learning*, secondo lo studio di Farazi (2024). Circa il settore dei prestiti, Becerra-Vicario *et al.* (2024) evidenziano inoltre come si possa avere un'analisi più specifica dei parametri dei mutuatari e di ridurre i rischi di insolvenza. Inoltre, le soluzioni basate sull'IA migliorano la sicurezza e l'efficienza delle reti di pagamento, grazie al *routing* intelligente delle transazioni e alla valutazione in tempo reale del rischio di frode. Anche su questo aspetto si esprime lo studio di Farazi (2024), secondo cui sarebbe possibile l'Identificazione delle transazioni fraudolente attraverso tecniche di rilevamento delle anomalie consentito dal *machine learning*.

Il paper di Jagdale e Deshmukh (2025) evidenzia che nell'ambito finanziario la NLP è diventata una forza trasformativa. Il valore aggiunto risiederebbe nel fatto che tale strumento concede alle aziende la possibilità di estrarre preziose informazioni da fonti di dati non strutturati come articoli di notizie, rapporti sugli utili e feed dei social media. Tale capacità sarebbe cruciale per prendere decisioni finanziarie informate. Le metodologie fondamentali della NLP su tale particolare settore, secondo il paper di Jagdale e Deshmukh (2025), sono innanzitutto la tokenizzazione, la quale permetterebbe la suddivisione del testo in unità più piccole per l'analisi. Inoltre, si annovera la sentiment analysis, cioè la valutazione del sentimento espresso nei testi finanziari per misurare il sentiment del mercato. In terza battuta si evince la modellazione dei temi, cioè l'identificazione di temi e argomenti all'interno di grandi dataset. Il paper esamina poi le varie applicazioni della NLP in finanza, tra cui il trading algoritmico. Tale strumento permetterebbe di utilizzare la NLP per analizzare il sentiment del mercato e prendere decisioni di trading. Successivamente, si evince che l'IA supporterebbe su un'altra importante attività: lo sfruttamento delle informazioni dai dati non strutturati per valutare l'affidabilità creditizia, nonché l'identificazione di attività fraudolente attraverso l'analisi dei dati testuali.

Infine, modelli NLP avanzati come i *transformer* e gli *embeddings*, hanno il potenziale di migliorare significativamente l'analisi dei dati finanziari. Questi modelli possono elaborare grandi volumi di dati in modo più efficace, portando a migliori intuizioni e previsioni.

Tenendo conto dei significativi cambiamenti dovuti alle innovazioni tecnologiche, diversi

studi confermano che l'integrazione del fintech e dell'Intelligenza artificiale nel settore bancario riduce la domanda per ruoli tradizionali, come funzionari di prestito e consulenti, promuovendo processi di disintermediazione finanziaria (Duygun *et al.* 2021*;* Jiang *et al.* 2021). Parallelamente, cresce la richiesta di figure professionali con competenze ibride, che combinano finanza, analisi dei dati, cybersecurity e intelligenza artificiale. Per mitigare i rischi occupazionali e favorire l'adattamento, diventano essenziali programmi di riqualificazione e aggiornamento continuo (Jiang *et al.* 2021).

Alla luce di questa analisi che evidenzia una significativa implementazione di strumenti di IA ben individuati nel paragrafo, è utile comprendere l'impatto dei lavoratori del settore bancario finanziario e assicurativo al fine di indagare sui trend occupazionali che caratterizzano il settore e che potrebbero essere stati influenzati dall'IA. In tal senso si propone di utilizzare, innanzitutto l'indice di esposizione delle professioni all'Intelligenza artificiale AIOE (Felten *et al.* 2021) applicato al contesto italiano (Ferri *et al.* 2024) e successivamente l'Indice C-AIOE (Pizzinelli *et al.* 2023) applicato al contesto italiano (Ferri *et al.* 2024).

A seguito di questa rassegna di letteratura è utile capire se la alta esposizione dei servizi finanziari e assicurativi, testimoniata dalle numerose innovazioni sopra descritte, corrisponde a un rallentamento delle attivazioni di contratti oppure questa forte esposizione non abbia alcun tipo di effetto.

2. Dati e metodologia

Dopo la prima rassegna di letteratura e un approfondimento molto specifico sui sistemi d'implementazione di IA disponibili nel settore bancario e assicurativo, il lavoro intende affrontare in prima battuta il quadro descrittivo e di contesto sull'occupazione dell'IA nelle aziende italiane. Successivamente, il lavoro si propone di applicare due indici noti in letteratura e sviluppati negli studi precedenti su delle prime analisi descrittive e su una serie di banche dati che ci permettano di leggere i dati relativi all'esposizione all'IA dei lavoratori focalizzando l'attenzione soprattutto sui servizi bancari, finanziari e assicurativi.

Il primo Indice che si utilizza è stato sviluppato da Felten *et al.* (2021). Per le 10 applicazioni individuate come rilevanti per l'IA sono stati dunque raccolti 52 valori che possono essere associati a ciascuna delle attitudini presenti nel sistema di Classificazione delle Professioni italiana, equivalenti alle *abilities* del sistema statunitense O*NET. Utilizzando questi valori, è stata calcolata l'esposizione all'IA per ogni occupazione tramite la formula dell'AIOE. Questo passaggio è stato realizzato impiegando i dati dell'Indagine campionaria delle Professioni del 2013 anziché quelli dell'O*NET 2020, ritenendo che ciò possa superare le critiche relative alla costruzione dell'AIOE basata sul contesto statunitense. A tal proposito sono stati considerati al nominatore livello, importanza ed esposizione all'IA per ogni abilità e al denominatore livello e importanza (Ferri *et al.* 2024).

$$AIOE_K = \frac{\sum_{j=1}^{52} A_{ij}\, x L_{jk} x I_{jk}}{\sum_{j=1}^{52} L_{jk} x I_{jk}}$$

Successivamente, teniamo conto di ciò che Pizzinelli *et al.* (2023) chiamano "*work context*" and "*job zones*". Le prime sono tradotte in Italia come "condizioni di lavoro", sono 57 e ne vengono selezionate 11. Le altre riguardano il livello d'istruzione.

Le 6 componenti identificate sono le seguenti:

1. Comunicazione – i) Face to face (H1) ii) Public speaking (H2);
2. Responsabilità – i) Responsabilità per I risultati (H11) ii) Responsabilità per la salute degli altri (H10);
3. Condizioni fisiche – i) Esposizione all'ambiente esterno (H17) ii) prossimità fisica agli altri (H21);
4. Criticità – i) Conseguenze dei propri errori (H45) ii) Libertà delle decisioni (H48) iii) Frequenza delle decisioni (H47);
5. Routine – i) Grado di automazione (H49) ii) Lavoro strutturato o non strutturato (H56);
6. *Skills* – i) *Job zones* (Forze di lavoro).

La formula è di seguito riportata:

$$C - AIOE_i = AIOE_i * (1 - (\boldsymbol{\theta}_1\text{-}\boldsymbol{\theta}_{MIN})) \quad (3)$$

Questi indici sono innanzitutto utilizzati per realizzare delle analisi descrittive sulle banche dati a disposizione dell'Inapp. Successivamente, si utilizzeranno per realizzare delle stime che permettano di comprendere come il settore in questione potrebbe essere influenzato dal punto di vista occupazionale.

Nello specifico in questa fase in cui gli indici sono ancora in una fase di test è possibile applicarli su vari

sistemi classificatori e banche dati per capire se la prima posizione occupata dai servizi finanziari e assicurativi sia effettivamente coerente sia nella lettura degli Ateco, sia nella lettura delle professioni (CP) che ricadono nei Settori economico-professionali di Atlante[3].

L'analisi successiva si concentrerà sulle transizioni professionali tra diverse occupazioni, per comprendere dove si siano spostate, a partire dal 2015, le persone impiegate nei servizi finanziari e assicurativi, secondo i dati SISCO. Esaminare questi percorsi consente di valutare se vi siano stati passaggi verso ruoli più o meno esposti rispetto a quelli precedenti. L'analisi sarà arricchita da evidenze empiriche utili a cogliere i primi effetti occupazionali legati alla crescente esposizione di questo comparto all'innovazione tecnologica.

Le analisi econometriche si propongono di stimare una variabile dipendente relativa alla variazione media annua percentuale di attivazione dei contratti nell'ambito di ogni professione a livello di macroaree italiane. S'intende comprendere se far capo a una professione dei servizi finanziari e assicurativi cambi la variazione percentuale media di attivazione dei contratti. Il modello di regressione lineare prevede la seguente equazione di riferimento:

$$\Delta y_{i,k} = \beta_0 + \beta_1 \mathrm{IA}_k + \beta_2 \mathrm{Settore}_k + \beta_3 (\mathrm{IA_k} \times \mathrm{Settore_k}) + \beta_4 + \varepsilon_{i,k} \quad (4)$$

$$\Delta y_{i,k} = \beta_0 + \beta_1 \mathrm{IA}_k + \beta_2 \mathrm{Settore}_k + \beta_3 (\mathrm{IA}_k \times \mathrm{Settore}_k) + \beta_4 + \beta_5 (\mathrm{IA}_k \times \mathrm{Routine}_k) + \varepsilon_{i,k} \quad (5)$$

Nella formula (4) y è la variazione percentuale media annua dei contratti attivati dal 2018 al 2023 per macroarea i (Nord, Centro e Sud) e occupazione k; IA rappresenta l'esposizione all'Intelligenza artificiale della professione k (utilizziamo alternativamente l'Indice CIAOE oppure l'Indice AIOE), Settore è la variabile che indica se la professione identificata è quella relativa alle Banche, finanza e assicurazioni (0/1); $\mathrm{IA_k} \times \mathrm{Settore_k}$ è l'interazione tra l'esposizione all'IA e il settore stesso.

La variabile successiva comprende un insieme di caratteristiche osservabili della forza lavoro iniziale nel 2018 (quota di donne, stranieri, laureati, part-time, over 50, retribuzione media, ore medie lavorate, contratti a tempo indeterminato), incluse per controllare possibili differenze strutturali tra le professioni nelle diverse aree del Paese.

Nella formula (5), viene inclusa un'ulteriore interazione tra l'esposizione all'IA (AIOE) e la routinarietà della professione $\mathrm{IA_k} \times \mathrm{Routine_k}$, per testare se l'effetto dell'IA sia più pronunciato nelle occupazioni caratterizzate da mansioni routinarie.

Tale variabile viene inclusa nelle stime che considerano l'AIOE (%), il quale non integra la routinarietà, e per questo è possibile isolarne l'effetto.

Gli errori standard sono clusterizzati a livello di combinazione macroarea-professione per tener conto della possibile eteroschedasticità e della dipendenza intra-cluster. La scelta di focalizzare l'analisi sulla variazione percentuale media annua delle attivazioni di contratto, a livello di macroarea e per il periodo 2018-2023, nasce dall'esigenza di cogliere eventuali segnali di cambiamento strutturale nella domanda di lavoro, potenzialmente legati alla crescente diffusione dell'Intelligenza artificiale. Tale misurazione permette, pertanto, di sintetizzare in modo efficace le dinamiche occupazionali, filtrando le fluttuazioni congiunturali di breve periodo e permettendo di evidenziare trend sistemici che potrebbero essere associati all'adozione di nuove tecnologie. In tale contesto, l'inclusione di una variabile dummy per il settore dei servizi finanziari e assicurativi risponde all'esigenza di isolare un comparto storicamente tra i primi ad adottare soluzioni di IA per attività routinarie e cognitive (come l'elaborazione dati, la gestione del rischio o il customer care automatizzato). La scelta di includere l'Indice AIOE e CAIOE come proxy dell'esposizione all'IA permette di verificare come l'impatto dell'IA sul mercato del lavoro si manifesti.

Nel disegno empirico adottato, si è scelto di concentrare l'analisi sulle attivazioni di contratto e non sulle interruzioni, per diverse ragioni di ordine teorico e metodologico. Anzitutto si possono cogliere in tal modo segnali di trasformazione nella domanda di lavoro potenzialmente indotti dalla diffusione dell'Intelligenza artificiale. In questo senso, le attivazioni rappresentano una misura più diretta dei nuovi ingressi nel mercato del lavoro o della volontà delle imprese di assumere. Dal

3 Di seguito si inseriscono le professioni comprese nel SEP dell'Atlante del Lavoro e delle qualificazioni 3.3.2.2.0 - Tecnici del lavoro bancario; 3.3.2.3.0 - Agenti assicurativi; 3.3.2.4.0 - Periti, valutatori di rischio e liquidatori; 3.3.2.5.0 - Agenti di borsa e cambio, tecnici dell'intermediazione titoli e professioni assimilate; 3.3.2.6.1 - Tecnici dei contratti di scambio, a premi e del recupero crediti; 3.3.2.6.2 - Tecnici della locazione finanziaria; 4.2.1.1.0 - Addetti agli sportelli assicurativi, bancari e di altri intermediari finanziari; 4.2.1.4.0 - Addetti agli sportelli delle agenzie di pegno e professioni assimilate.

punto di vista empirico, le attivazioni risultano più stabili e comparabili nel tempo e tra territori, soprattutto se misurate in termini di variazione percentuale media annua.

D'altra parte, anche il numero di attivazioni di contratto può essere influenzato da una pluralità di fattori – come dinamiche congiunturali, politiche attive, incentivi o comportamenti delle imprese. La scelta di utilizzare la variazione percentuale media annua delle attivazioni, tuttavia, consente di mitigare parte di questa eterogeneità.

Inoltre, rispetto alle interruzioni, le attivazioni potrebbero essere maggiormente collegate alle scelte di investimento produttivo e alla domanda di nuove competenze da parte delle imprese e potrebbero rappresentare una variabile più adatta per indagare gli effetti dell'esposizione all'IA. Una nuova attivazione rappresenta di fatto una decisione proattiva dell'impresa che, in presenza di tecnologie sostitutive o complementari, potrebbe risultare condizionata proprio dal grado di esposizione delle professioni all'IA.

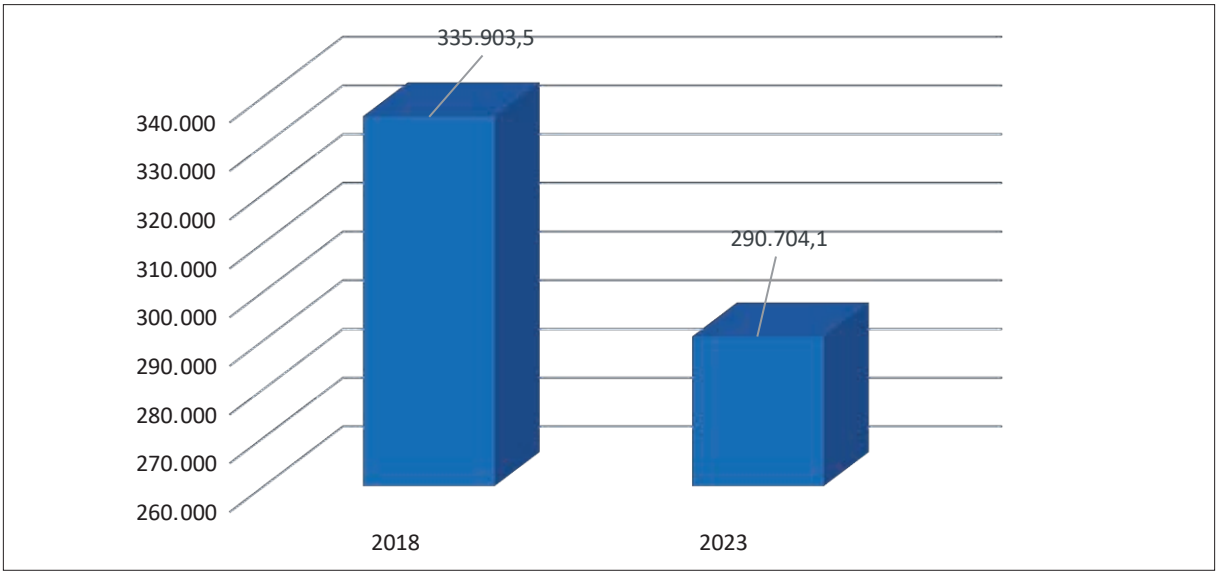
3. Analisi descrittive. Come cambia negli anni più recenti l'occupazione nei servizi bancari, finanziari e assicurativi

Per comprendere come è cambiata l'occupazione nel settore dei servizi finanziari e assicurativi negli ultimi anni, si utilizza anzitutto l'Indagine campionaria delle Forze di lavoro, unendo per la prima volta i descrittivi di Atlante del Lavoro e delle qualificazioni che permettono di osservare i profili professionali più coerenti relativi ai lavori assicurativi, finanziari e bancari. Dal 2018 al 2023 il settore economico professionale relativo ai servizi finanziari e assicurativi ha perso il 13,5% degli occupati.

Classificando i dati della Rilevazione delle Forze di lavoro con i SEP di Atlante, possiamo osservare che l'indice risulta mediamente molto elevato, mostrando quindi che gli occupati nel settore economico-professionale sono fortemente esposti all'IA; sembrerebbero meno esposti i SEP dell'edilizia e della agricoltura, silvicoltura e pesca.

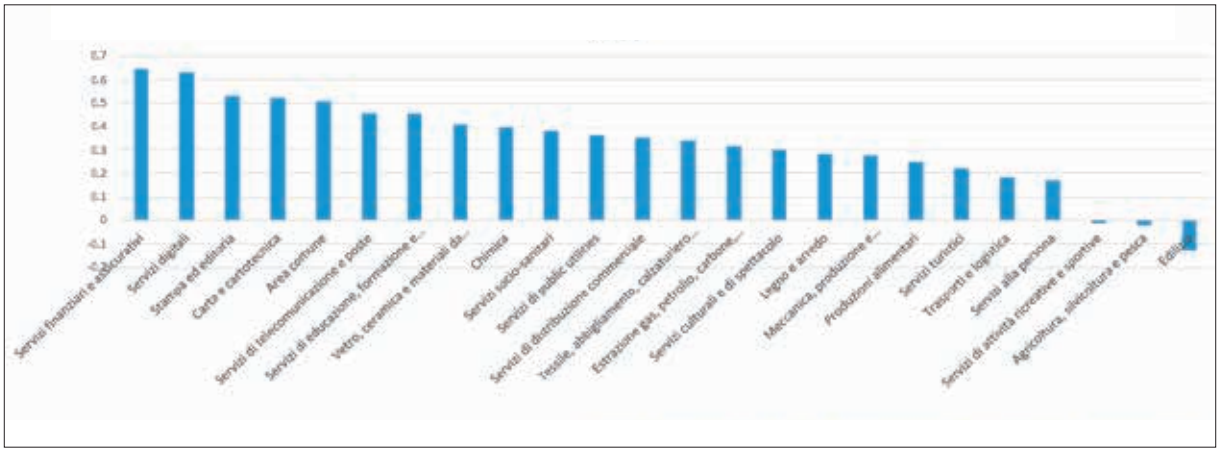
Osservando i dati della Rilevazione Forze di lavoro rispetto ai settori Ateco, emerge che nell'ambito delle attività finanziarie e assicurative si rileva la maggiore concentrazione di lavoratori potenzialmente sostituibili, il 91,7% dei lavoratori occupati nel settore fa parte della categoria dei potenzialmente sostituibili. In netta contrapposizione con il settore dell'agricoltura, silvicoltura e pesca in cui tale percentuale corrisponde al 3,1% e invece la quota dei non sostituibili è dell'82,62%.

Figura 1. Occupati nel settore economico professionale dei servizi bancari, finanziari e assicurativi



Fonte: elaborazione degli Autori sulla Rilevazione continua sulle Forze di lavoro 2018 e 2023

Figura 2. CAIOE - Potenziale esposizione delle professioni all'IA rispetto ai Settori economico-professionali



Fonte: elaborazione degli Autori su dati Atlante del Lavoro e delle qualificazioni

Tabella 1. Percentuale di lavoratori per gruppo creato sulla base dell'Indice CAIOE rispetto ai 12 settori Ateco aggregati (dati di stock)

	Agricoltori, silvicoltori	Industria in senso stretto	Costruzioni	Commercio	Alberghi e ristoranti	Trasporto e magazzinaggio	Servizi di informazione e comunicazione	Attività finanziarie	Attività immobiliari	Amministrazione pubblica	Istruzione, sanità	Altri servizi collettivi	Totale
Potenzialmente non sostituibili	82,62	17,99	57,45	13,93	52,88	37,77	1,46	0,38	16,96	12,69	6,97	47,25	24,82
Non sostituibili ma coadiuvati su alcuni tasks	12,06	31,53	23,12	47,51	32,68	21,08	1,98	0,34	8,71	10,82	37,41	27,32	27,15
Coadiuvati con alcuni tasks sostituibili	2,23	31,71	10,83	20,16	11,63	12,73	24,05	7,59	28,32	12,46	36,2	13,28	22,55
Potenzialmente sostituibili	3,1	18,77	8,59	18,41	2,81	28,42	72,51	91,69	46,02	64,03	19,42	12,16	25,48
Totale	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**

Fonte: elaborazione degli Autori su dati Rilevazione continua sulle Forze di lavoro 2022

Prendendo in considerazione gli stessi anni di riferimento, cioè il 2018 come anno base, non influenzato dal Covid, che ha avuto un grande impatto sull'adozione di sistemi tecnologici avanzati e sull'occupazione, e il 2023 come anno finale, si evince che sul totale di 35.066,04 occupati in meno, alcune attività hanno perso un maggior numero di lavoratori. Il dato precedente che ha permesso di restringere il campo, grazie all'Atlante del Lavoro e delle qualificazioni, evidenzia una perdita di occupazione molto più rilevante proprio perché con i settori Ateco s'include l'occupazione di qualsiasi professione lavori all'interno di un'azienda catalogata in un determinato modo.

Nell'ambito delle attività ausiliarie dei servizi finanziari e delle attività assicurative, l'occupazione invece sembrerebbe in aumento.

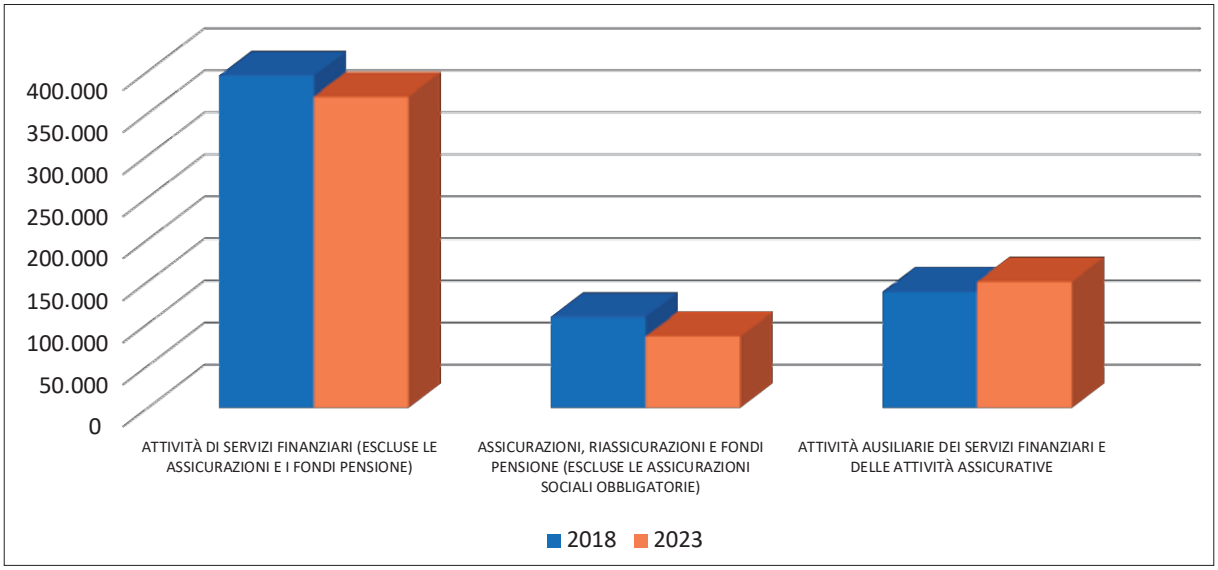
Dal 2018 al 2023, il settore dei servizi finanziari ha perso il 6,25% degli occupati, mentre il settore delle assicurazioni ha visto una riduzione del 21,06%. Le attività ausiliarie dei servizi finanziari, diversamente, hanno registrato un aumento del 9,01%. Pertanto, il numero totale di occupati nei settori finanziari e assicurativi è diminuito del 5,46%, riflettendo una trasformazione del mercato del lavoro dovuta all'adozione di tecnologie avanzate.

Tabella 2. Occupati nei tre settori Ateco a due digit 64, 65 e 66 che compongono i servizi finanziari e assicurativi

Settori Ateco	2018	2023	Variazione % (2023 su 2018)
64. Attività di servizi finanziari	395.263,60	370.568,83	**-6,25%**
65. Assicurazioni, riassicurazioni e fondi pensione	108.335,20	85.512,10	**-21,06%**
66. Attività ausiliarie dei servizi finanziari	138.154,90	150.606,70	**+9,01%**
Totale	**641.753,70**	**606.687,63**	**-5,47%**

Fonte: Rilevazione continua Forze di lavoro 2018-2023

Figura 3. Occupazione nell'ambito dei settori Ateco dei servizi finanziari e assicurativi



Fonte: elaborazione degli Autori sulla Rilevazione continua sulle Forze di lavoro 2018 e 2023

Secondo i dati di flusso che si rifanno alle Comunicazioni Obbligatorie, dopo la battuta d'arresto dovuta al periodo pandemico, si rileva un aumento che porta al superamento dei livelli del 2018 delle attivazioni di contratto negli Ateco oggetto d'interesse; l'aumento calcolato è del +14,38% dal 2018.

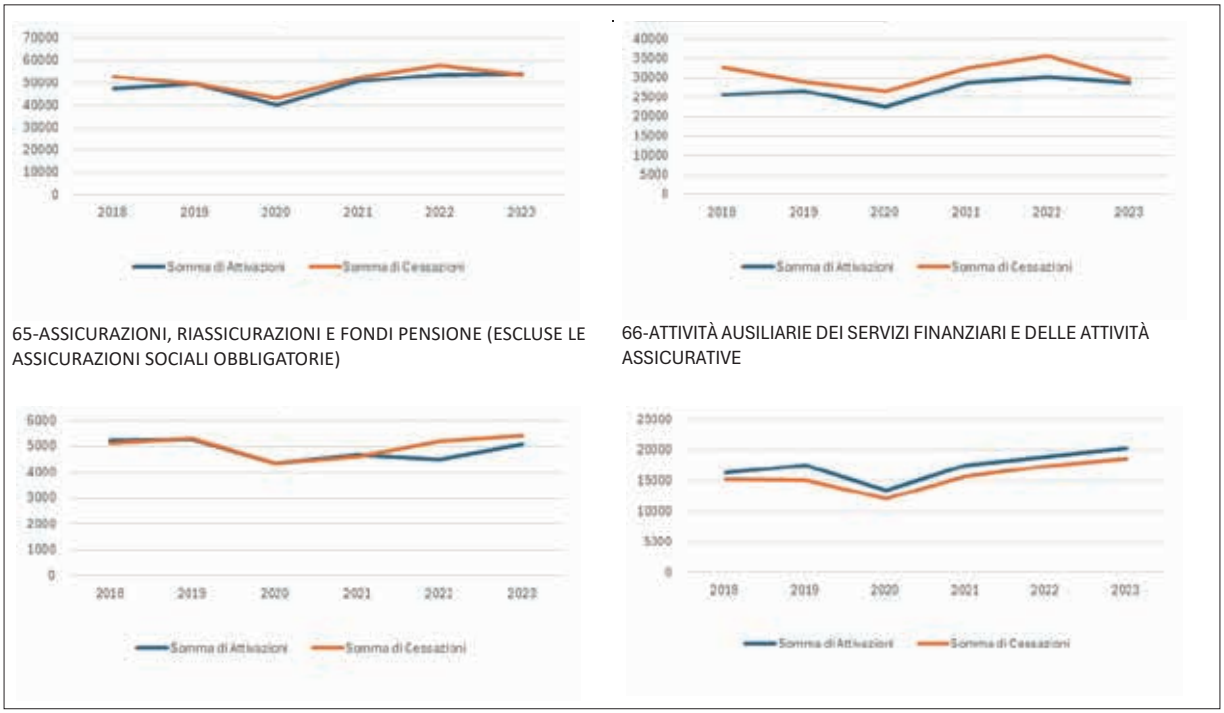
Nel settore Ateco 64 relativo alle Attività dei servizi finanziari, la variazione percentuale di +11,8% indica un aumento delle attivazioni di contratto in questo settore dal 2018 al 2023. Il differenziale percentuale medio annuo è del 3,21%.

La variazione percentuale nel settore delle assicurazioni, riassicurazioni e fondi pensione (escluse le assicurazioni sociali obbligatorie) di -2,7% suggerisce una leggera riduzione dell'occupazione dal 2018 al 2023. Il differenziale percentuale medio annuo è dello 0,01%.

Nel settore delle attività ausiliarie dei servizi finanziari e delle attività assicurative si evince un aumento del 24% dal 2018 al 2023. Il differenziale percentuale medio annuo è del 5,87%.

Leggendo gli stessi dati attraverso i settori economico-professionali di Atlante che restringono il campo alle sole professioni coerenti con il settore bancario finanziario e assicurativo, si evince un aumento del 12,35% delle attivazioni nei profili più coerenti. Osserviamo le professioni più significative in termini numerici e il loro andamento: circa i tecnici del lavoro bancario si evince un differenziale negativo, questo potrebbe suggerire una riduzione complessiva dell'occupazione in questo settore, che peraltro è quella più influente sul totale complessivo. Il totale delle attivazioni di contratto dal 2018 è di 45.302,42, invece le cessazioni sono state negli stessi anni 73.984,49. La differenza tra cessazioni e attivazioni è di -28682,07, numero molto elevato che farebbe il paio con le evidenze di letteratura che raccontano di un settore con una serie di difficoltà occupazionali legate all'ascesa delle nuove tecnologie.

Figura 4. Attivazioni e cessazioni dalle Comunicazioni Obbligatorie nell'ambito dei settori Ateco dei servizi finanziari e assicurativi



Fonte: elaborazione degli Autori su dati SISCO dal 2018 al 2023

Per quanto riguarda gli Agenti assicurativi, in tutti gli anni considerati, le attivazioni superano le cessazioni. Continuando sempre ad analizzare le professioni che quotano un maggior numero di lavoratori (ed escludendo in questa fase dalle analisi le professioni meno significative in termini di occupati) nell'ambito della professione Tecnici dei contratti di scambio, a premi e del recupero crediti, si evince un differenziale sempre positivo, indicando che le attivazioni superano le cessazioni, eccetto nell'ultimo anno in cui le due linee tendono ad avvicinarsi.

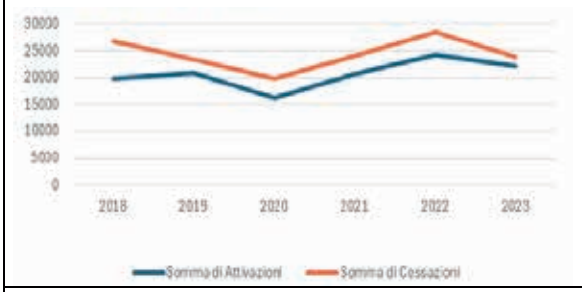
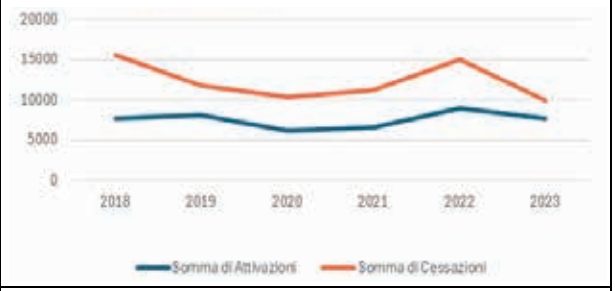
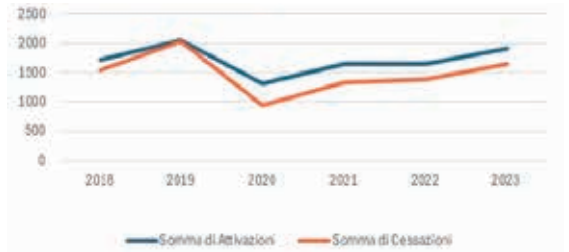
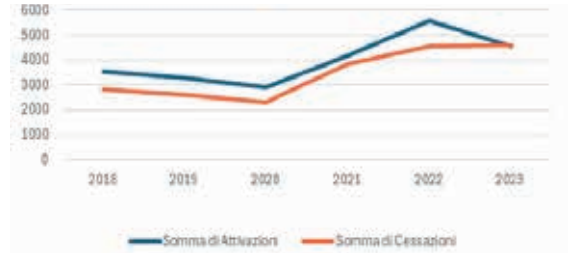
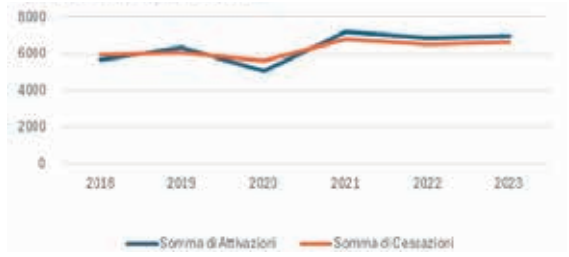
La tabella di seguito è utile a comprendere come siano state le trasformazioni professionali dei lavoratori che fanno capo alle professioni del SEP di Atlante inerente ai servizi finanziari e assicurativi. I Tecnici del lavoro bancario che hanno cambiato lavoro, sono diventati nell'ordine Addetti agli affari generali, Addetti agli sportelli assicurativi, bancari e di altri intermediari finanziari, Tecnici della gestione finanziaria, Specialisti in risorse umane, Esercenti di attività sportive, Specialisti in attività finanziarie e Analisti e progettisti di software ecc. Dal 2015 ad oggi la professione più prossima è stata, pertanto, quella degli Affari generali, con il 20,27% dei lavoratori che si spostano in questa categoria. Questo potrebbe riflettere una vicinanza delle competenze amministrative generali, ragione per cui si potrebbe giustificare tale cambiamento.

L'8,44% si sposta verso la professione degli Addetti agli sportelli assicurativi, bancari e di altri intermediari finanziari e il 5,66% passa ai Tecnici della gestione finanziaria, mantenendo continuità e specializzazione nel settore finanziario.

Successivamente, nell'ordine, il 26,38% degli Addetti agli sportelli assicurativi si spostano principalmente verso la professione Addetti agli affari generali, suggerendo una versatilità nelle competenze amministrative. Il 9,76% mostra una mobilità interna al settore bancario e diventa Tecnico del lavoro bancario. Quest'ultima professione, appartenente al grande gruppo tre dei tecnici, potrebbe anche essere identificata come una progressione ed evoluzione nella carriera dell'individuo, che parte da Addetto, facente capo al grande gruppo quattro.

Gli Agenti assicurativi confermano come in tutti gli altri casi considerati la transizione prevalente verso Addetti agli affari generali. Il 5,71% dei lavoratori si è spostato invece verso la professione

Figura 5. Occupazione dalle Comunicazioni Obbligatorie nell'ambito delle professioni

TOTALE	**3.3.2.2.0 - Tecnici del lavoro bancario**
	
3.3.2.3.0 - Agenti assicurativi	**3.3.2.6.1 - Tecnici dei contratti di scambio, a premi e del recupero crediti**
	
4.2.1.1.0 - Addetti agli sportelli assicurativi, bancari e di altri intermediari finanziari	
	

Fonte: elaborazione degli Autori su dati SISCO dal 2018 al 2023

dei Commessi alle vendite al minuto, suggerendo una mobilità verso ruoli di vendita. Altre opzioni includono il passaggio alla professione di Addetti agli sportelli assicurativi, bancari e di altri intermediari finanziari con una percentuale del 4,44%.

Circa la professione di Periti, valutatori di rischio, liquidatori e professioni assimilate, oltre la solita prima tipologia di spostamento verso Addetti agli affari generali, si annovera lo spostamento dell'11,19% dei lavoratori che diventano Agente assicurativo, mostrando una mobilità interna nel settore assicurativo.

Altre transizioni includono Addetti all'informazione e all'assistenza dei clienti, Addetti agli sportelli assicurativi, bancari e di altri intermediari finanziari e Tecnici della produzione manifatturiera, con percentuali rispettivamente del 3,97%, 3,09% e 2,54%.

I Tecnici dei contratti di scambio, a premi e del recupero crediti, dopo la transizione principale verso gli Addetti agli affari generali, con il 12,01% dei lavoratori, si spostano su altri percorsi professionali che Venditori a distanza, Addetti all'informazione nei Call center, Addetti agli sportelli per l'esazione di imposte e contributi e al recupero crediti e Commessi delle vendite al minuto, con percentuali rispettivamente del 3,77%, 3,52% 2,84% e 2,49%.

Tabella 3. Percentuale delle transizioni di contratto dalla professione originaria (33220 - Tecnici del lavoro bancario ad altra professione) 2015-2023

Tecnici del lavoro bancario (Codice cp: 33220)		
3.3.2.2.0 → 4.1.1.2.0	Addetti agli affari generali	20,27
3.3.2.2.0 → 42110	Addetti agli sportelli assicurativi, bancari e di altri intermediari finanziari	8,44
3.3.2.2.0 → 33210	Tecnici della gestione finanziaria	5,66
3.3.2.2.0 → 25131	Specialisti in risorse umane	3,89
3.3.2.2.0 → 54214	Esercenti di attività sportive	2,79
Addetti agli sportelli assicurativi, bancari e di altri intermediari finanziari (Codice cp: 42110)		
4.2.1.1.0 → 4.1.1.2.0	Addetti agli affari generali	26,38
4.2.1.1.0 → 3.3.2.2.0	Tecnici del lavoro bancario	9,76
4.2.1.1.0 → 4.1.1.1.0	Addetti a funzioni di segreteria	4,4
4.2.1.1.0 → 5.1.2.2.0	Commessi alle vendite al minuto	3,85
4.2.1.1.0 → 5.4.2.1.4	Esercenti di attività sportive	2,01
Agenti assicurativi (Codice cp: 33230)		
3.3.2.3.0 → 4.1.1.2.0	Addetti agli affari generali	27,06
3.3.2.3.0 → 5.1.2.2.0	Commessi alle vendite al minuto	5,71
3.3.2.3.0 → 4.2.1.1.0	Addetti agli sportelli assicurativi, bancari e di altri intermediari finanziari	4,44
3.3.2.3.0 → 4.1.1.1.0	Addetti a funzioni di segreteria	3,23
3.3.2.3.0 → 4.2.2.4.0	Addetti all'informazione nei Call Center (senza funzioni di vendita)	2,94
Periti, valutatori di rischio, liquidatori e professioni assimilate (Codice cp: 33240)		
3.3.2.4.0 → 4.1.1.2.0	Addetti agli affari generali	27,84
3.3.2.4.0 → 3.3.2.3.0	Agenti assicurativi	11,19
3.3.2.4.0 → 5.1.3.4.0	Addetti all'informazione e all'assistenza dei clienti	3,97
3.3.2.4.0 → 4.2.1.1.0	Addetti agli sportelli assicurativi, bancari e di altri intermediari finanziari	3,09
3.3.2.4.0 → 3.1.5.3.0	Tecnici della produzione manifatturiera	2,54
Tecnici dei contratti di scambio, a premi e del recupero crediti (Codice cp: 33261)		
3.3.2.6.1 → 4.1.1.2.0	Addetti agli affari generali	12,01
3.3.2.6.1 → 5.1.2.5.2	Venditori a distanza	3,77
3.3.2.6.1 → 4.2.2.4.0	Addetti all'informazione nei Call Center (senza funzioni di vendita)	3,52
3.3.2.6.1 → 4.2.1.3.0	Addetti agli sportelli per l'esazione di imposte e contributi e al recupero crediti	2,84
3.3.2.6.1 → 5.1.2.2.0	Commessi delle vendite al minuto	2,49

Nota: Sono incluse qui solo le professioni con un numero più elevato di 5.000 transizioni dal 2015 in poi.

Fonte: elaborazione degli Autori su dati SISCO

Va osservato che l'Indice di esposizione all'IA (CAIOE) stimato per i Tecnici del lavoro bancario è tra i più alti (0,711909) e anche la professione degli Addetti agli affari generali presenta un livello di esposizione quasi identico (0,711759). Questo vuol dire che i lavoratori transitano principalmente da professioni molto esposte a professioni egualmente esposte e che probabilmente quanto osservato sui dati delle Comunicazioni Obbligatorie implica una possibile difficoltà futura qualora i possibili scenari negativi relativi allo spiazzamento dei lavoratori dovessero verificarsi. È opportuno osservare che gli Addetti agli sportelli assicurativi, bancari e di

altri intermediari finanziari sono collocati ancora più in alto nel gradiente di esposizione all'IA (0,736309).

Situazione diversa si osserva per i Commessi alle vendite al minuto in cui l'AIOE è poco più di 0,22, quindi un'esposizione molto più bassa, gli Agenti assicurativi invece hanno un indice di 0,629684. I Venditori a distanza hanno un'esposizione ancor più elevata (0,768562) e sarebbero la professione più vicina in cui transitano i Tecnici di contratti di scambio, che sarebbero peraltro meno esposti (0,620536).

In conclusione, quel che emerge dalla tabella 3 è che la mobilità da una professione all'altra è stata negli ultimi anni sicuramente verso professioni che fanno capo sempre all'ambito bancario, nonché in Addetti agli affari generali con cui hanno in comune competenze amministrative. Anche quando lo spostamento è verso professioni diverse rispetto al settore finanziario assicurativo, l'esposizione della professione resta molto elevata. Quando c'è mobilità, pertanto, questa avviene all'interno dello stesso macro-ambito professionale o in direzioni che ne condividono le competenze cardine, come quelle amministrative o organizzative. In altre parole, la professionalità maturata nel settore finanziario-assicurativo tende a restare 'spendibile' in percorsi affini, suggerendo che il capitale umano sviluppato in questi ambiti è ancora ampiamente valorizzato ma potrebbe avere in un futuro maggiori difficoltà nel transitare da una professione all'altra qualora si rendesse necessario un percorso di *reskilling*.

3. Analisi econometriche

Nelle analisi econometriche di seguito presentate si vuole comprendere se il cambiamento nelle attivazioni di contratto possa avere un legame con la maggiore esposizione all'IA (tabella 4). Si intende verificare se il cambiamento nelle attivazioni contrattuali possa essere associato a una maggiore esposizione delle professioni all'IA. Le stime presentate nella tabella 4 si concentrano sulla variabile dipendente, ovvero la variazione percentuale media annua delle attivazioni di contratto registrata a livello di macroarea e per professione nel periodo 2018-2023. Tali dati sono stati elaborati utilizzando la banca dati SISCO dei datori di lavoro.

Le variabili indipendenti includono, nella prima colonna, l'Indice CAIOE che misura l'esposizione all'IA e una dummy relativa all'appartenenza a una professione nel settore dei servizi finanziari e assicurativi. I risultati mostrano un effetto negativo costante associato alla variabile che identifica le professioni nel settore dei servizi finanziari e assicurativi. In particolare, nelle prime tre specificazioni del modello, i coefficienti stimati indicano che, a parità delle altre condizioni, l'appartenenza a questo settore è associata a una riduzione media della variazione percentuale annua delle attivazioni contrattuali pari, rispettivamente, a -7,3 p.p., -19,8 p.p. e -21,4 p.p.

Questo risultato suggerisce che le professioni del comparto finanziario e assicurativo — notoriamente più esposte ai processi di automazione legati all'Intelligenza artificiale — stanno sperimentando un rallentamento nella crescita delle attivazioni dei contratti. Tali effetti possono essere interpretati come segnali di sostituzione tecnologica o di riorganizzazione interna delle attività, dove alcune funzioni tradizionali vengono automatizzate, riducendo la necessità di nuove assunzioni.

Nella seconda colonna, si aggiungono le caratteristiche della forza lavoro: entrambi i coefficienti (CAIOE e settore) risultano negativi e significativi, suggerendo che sia l'esposizione all'IA sia l'appartenenza al settore bancario e assicurativo abbiano contribuito alla riduzione delle attivazioni contrattuali nel periodo osservato. Nella terza colonna, l'aggiunta della variabile interagita elide la significatività dei coefficienti e altera la stabilità del modello. Verificando i test di collinearità (VIF), infatti, la variabile interagita crea molti problemi alla stima e il settore maggiormente collineare è proprio quello di nostro interesse. Questo risultato suggerisce che l'interazione tende ad assorbire parte dell'effetto settoriale, rendendo difficile distinguere empiricamente tra l'impatto dell'esposizione tecnologica e quello della specificità settoriale, data la loro forte correlazione strutturale. Per affrontare questo problema, si conclude, quindi, con la quarta colonna, in cui si mantiene la variabile CAIOE e si elimina quella settoriale in quanto collineare con l'interagita (CAIOE*Settore). La variazione delle attivazioni di contratti sembra avere una correlazione negativa e significativa rispetto al settore servizi finanziari e assicurativi interagito con l'esposizione all'Intelligenza artificiale misurata attraverso il CAIOE. In altri termini, quanto maggiore è l'esposizione delle professioni finanziarie e assicurative all'Intelligenza artificiale, tanto più si

Tabella 4. Regressione OLS per stimare la variazione percentuale media annua di nuove attivazioni di contratto dal 2018 al 2023 (variabile d'interesse CAIOE)

	Variazione attiv. contratti (1)	Variazione attiv. contratti (2)	Variazione attiv. contratti (3)	Variazione attiv. contratti (4)
CAIOE	3,6492	-9,6213**	7,4553	16,9274
	[4,0315]	[4,8514]	[13,2365]	[18,9140]
Servizi finanziari e assicurativi	-7,3671	-19,8043***	-21,4329	
	[6,0932]	[7,2923]	[42,3512]	
CAIOE_ Servizi finanziari e assicurativi	-12,2785	-44,8710**		
	[65,7272]	[20,2120]		
Caratteristiche dei lavoratori	No	Sì	Sì	Sì
_cons	13,2934***	29,1880***	27,5759***	22,0051***
	[5,0710]	[8,6938]	[9,0194]	[6,6295]
N	2107	2071	2071	2071
r2	0,027	0,0495	0,0516	0,0443

Fonte: elaborazione degli Autori su dati SISCO dal 2018 al 2023

osserva una riduzione della variazione percentuale annua nelle attivazioni di contratti.

Il modello suggerisce quindi che, nelle macroregioni italiane, la crescente esposizione all'IA di alcune professioni — in particolare nei servizi finanziari e assicurativi — sia associata a un rallentamento delle nuove attivazioni contrattuali. I risultati delle stime supportano l'ipotesi secondo cui le professioni appartenenti a tale comparto stanno attraversando un processo di razionalizzazione, che si traduce in una contrazione delle attivazioni contrattuali. Questo effetto, rilevato in modo coerente nelle diverse specificazioni del modello, è interpretabile come segnale di una transizione tecnologica che, almeno nel breve-medio periodo, riduce il fabbisogno occupazionale attraverso l'automazione di funzioni precedentemente svolte da lavoratori.

Tale approccio contribuisce al dibattito sugli effetti occupazionali dell'IA, offrendo una lettura delle dinamiche in atto e fornendo un primo riscontro empirico sulle trasformazioni in atto nei comparti maggiormente esposti a tale 'rivoluzione tecnologica'.

Le stime nella tabella 5 permettono di osservare separatamente l'esposizione all'IA e la routinarietà, un elemento che, come emergeva dalla letteratura, non sempre si riesce a scorporare dall'esposizione all'IA. Tale componente che nelle analisi precedenti era inclusa nell'Indice CAIOE, in questo caso è considerata separatamente. Anche per queste stime, la prima regressione consta anzitutto delle tre variabili essenziali, successivamente vengono aggiunte le caratteristiche dei lavoratori. Dai risultati sembrerebbe che l'esposizione potenziale delle professioni all'IA abbia una relazione positiva con la variabile dipendente e che quindi sia correlata a una variazione percentuale positiva di attivazione di contratti. Dal 2018 al 2023 le attivazioni di contratti sono state più positive per le professioni maggiormente esposte. Nella colonna 3, tuttavia, viene aggiunta la variabile interagita, che però, come nel caso precedente, introduce una certa collinearità. Per questo motivo, nella quarta colonna si include la variabile interagita tra l'esposizione all'IA e i servizi finanziari e assicurativi. I risultati cambiano nelle ultime colonne dove il coefficiente maggiormente degno di nota è l'esposizione all'IA moltiplicata per i Servizi finanziari e assicurativi, quest'ultima inciderebbe negativamente sulla variazione percentuale media annua dei contratti.

Pertanto, da quest'ultima batteria di regressioni si evince che la variazione media dei contratti attivati diminuisce di gran lunga qualora la professione sia potenzialmente più esposta e faccia capo al settore dei servizi finanziari e assicurativi. I risultati potrebbero quindi testimoniare che, essendo tale settore uno dei più investiti dalle ondate tecnologiche precedenti all'IA, nonché dall'incalzante ascesa dell'IA, sembrerebbe identificarsi già un andamento negativo della variazione percentuale di attivazione dei contratti per le professioni con un più alto indice

Tabella 5. Regressione OLS per stimare la variazione percentuale media annua di nuove attivazioni di contratto dal 2018 al 2023. (variabile d'interesse AIOE e Routine index)

	Variazione attiv. contratti (1)	Variazione attiv. contratti (2)	Variazione attiv. contratti (3)	Variazione attiv. contratti (4)	Variazione attiv. contratti (5)
	b/se	b/se	b/se	b/se	b/se
AIOE	6,4863**	-4,9775	12,9987	12,6597	
	[3,1827]	[3,9090]	[11,0053]	[14,0933]	
Routine	0,1776*	0,1449	0,1884*	0,144	-0,0548
	[0,1041]	[0,1087]	[0,1144]	[0,1100]	[0,1483]
Servizi finanziari e assicurativi	-7,0594	-18,7492**	-44,5115		
	[6,3677]	[7,3846]	[78,9631]		
AIOE_ Servizi finanziari e assicurativi		15,422	-31,0695**	-37,2586***	
			[96,8655]	[15,2636]	[13,2278]
AIOEROUT				0,443	
					[0,3089]
Caratteristiche dei lavoratori	No	Sì	Sì	Sì	Sì
_cons	3,5774	22,0803**	18,7766**	15,7220**	25,3723***
	[8,3228]	[8,9788]	[9,1252]	[6,9852]	[8,8943]
N	2107	2071	2071	2071	2071
r2	0,029	0,05	0,0532	0,0466	0,0474

Fonte: elaborazione degli Autori su dati SISCO dal 2018 al 2023

di esposizione all'IA che fanno parte del settore dei servizi finanziari e assicurativi.

Conclusioni

Come si è osservato in letteratura, le tre fasi tipiche dell'introduzione di una nuova tecnologia nel mercato del lavoro prevedono inizialmente una sostituzione diretta dei lavori e dei compiti attualmente svolti dai lavoratori (effetto di sostituzione), seguita da un aumento complementare dei lavori e dei compiti necessari per utilizzare, gestire e supervisionare le nuove macchine (effetto di complementarità delle competenze). Il timore, tuttavia, è che nel caso dell'IA, il ritmo di adozione degli strumenti da parte delle imprese cresca così rapidamente da accelerare i tempi delle tre fasi e che si debba quanto prima iniziare a immaginare come transitare da una professione all'altra o, meglio, da gruppi di attività tipici di alcune professioni ad altri che potrebbero anche far capo ad altre professioni. Da questo punto di vista, l'*upskilling* e il *reskilling* diventeranno fondamentali. Le implementazioni di tali processi saranno utili per governare e programmare al meglio, garantendo una forza lavoro pronta e preparata ad affrontare le sfide più rilevanti. La rassegna della letteratura ha inoltre evidenziato che i lavori sono costituiti da una serie di *tasks*, e poiché alcuni di questi *tasks* potranno essere automatizzati, i profili di lavoro potranno cambiare. L'aggiunta di nuovi *tasks* e la modifica di quelli esistenti piuttosto che eliminare completamente una professione, ne cambierebbero tuttavia il volto (Ernst *et al.* 2019). Un approccio sempre più orientato a identificare le attività e a capire eventualmente l'evoluzione e la domanda di nuove competenze potrebbe essere, pertanto, molto utile e gli strumenti a disposizione dell'INAPP consentono di effettuare molti approfondimenti in tal senso.

Dal lavoro di analisi emerge poi che il settore dei Servizi bancari, finanziari e assicurativi, con qualsiasi tassonomia venga analizzato (settore Ateco o SEP di Atlante), risulta tra quelli più esposti all'IA. Si tratta, senz'altro, di un settore che potrebbe essere pronto a una possibile riorganizzazione di processi e adeguamento ai nuovi sistemi tecnologici proprio perché avvezzo a cambiamenti importanti che si sono affermati negli ultimi decenni. I cambiamenti dovuti ai sistemi IA nel settore finanziario e assicurativo stanno comportando una sostituzione del lavoro umano, traducendosi sia in un aumento delle fuoriuscite dei lavoratori da un settore con

minore necessità di forza lavoro, sia in una sempre minore crescita di attivazioni di contratti. Le analisi descrittive confermano questa ipotesi poiché hanno evidenziato che negli anni 2018-2023, sullo stock dei lavoratori nel settore specifico sono stati già persi ca. 35.000 lavoratori e, nel complesso, le ragioni potrebbero essere dovute a un rapporto tra assunzioni e cessazioni sbilanciate. Si osserva, infatti, per i Tecnici del lavoro bancario, professione che raccoglie il maggior numero di occupati, una tendenza ormai consolidata in cui le cessazioni superano di gran lunga le attivazioni in tutti gli anni presi in considerazione.

Le analisi econometriche confermano poi, utilizzando le metriche di esposizione delle professioni all'Intelligenza artificiale, che una maggiore esposizione all'IA diminuisce mediamente la variazione delle attivazioni di contratti. L'esposizione poi, interagita con il settore oggetto di studio, diminuisce anch'essa di alcuni punti percentuali tale variazione di attivazioni nelle macroaree italiane. Si evidenzia così un'incidenza negativa e significativa sull'andamento delle attivazioni dei contratti per i lavoratori che fanno capo a professioni bancarie, finanziarie o assicurative ritenute fortemente esposte all'IA.

D'altra parte, occorre evidenziare che, per avere chiari i potenziali scenari e gli effetti che potrebbero derivarne, è importante cogliere nello specifico i dettagli delle applicazioni maggiormente utilizzate nelle banche, istituzioni finanziarie e assicurazioni. Per questa ragione nel lavoro sono state evidenziate nel dettaglio le applicazioni che intervengono nelle attività tipiche dell'individuo e nei segmenti di lavoro in modo da poter comprendere attraverso un ulteriore affinamento di metodi il grado di eventuale sostituibilità del lavoratore all'interno dell'organizzazione.

Per fare questo tipo di analisi, occorre studiare in modo capillare nei vari settori quali applicazioni potrebbero essere sostitutive dell'individuo in alcuni *taks* e quali risulterebbero un semplice supporto. Questo lavoro, in tal senso costituisce una base che potrebbe essere consolidata ed evoluta in termini di approfondimento, nonché mutuata su altri settori Ateco o settori economico-professionali (Atlante del Lavoro). La duplice lettura, infatti, in questo studio è stata molto utile e ha permesso di comprendere come alcuni fenomeni possano essere analizzati da più punti di vista, restituendo risultati ancora più utili ai fini della programmazione di politiche, proprio grazie alla doppia lettura. Sulla base delle evidenze analizzate, si osserva che, a fronte di un andamento generale delle attivazioni occupazionali tendenzialmente crescente, il settore in esame registra una crescita più contenuta rispetto a molti altri comparti. Questo dato sollecita una riflessione urgente sull'opportunità di rafforzare, attraverso adeguate politiche pubbliche, gli strumenti a disposizione dei lavoratori del settore per affrontare le trasformazioni in atto. In particolare, risultano imprescindibili interventi mirati di *upskilling*, per garantire la piena padronanza delle nuove tecnologie richieste.

Tali politiche sono fondamentali per prevenire fenomeni di spiazzamento occupazionale e per evitare che si acuiscano divari territoriali, penalizzando le aree meno avanzate. Ciò assume particolare rilievo considerando che il settore è spesso costituito da gruppi nazionali e sovranazionali, i quali – a differenza di altri comparti più localizzati – possono presentare una maggiore uniformità negli investimenti tecnologici, ma non necessariamente negli impatti occupazionali generati.

Bibliografia

Acemoglu D., Restrepo P. (2017), Secular stagnation? The effect of aging on economic growth in the age of automation, *American Economic Review*, 107, n.5, pp.174-179

Battistoni A., Ferri V. (2024), Il mercato del lavoro cambia pelle, *Il Libero Professionista Reladed*, n.28, 7 novembre

Becerra-Vicario R., Salas-Compás B., Valcarce-Ruiz L., Sánchez-Serrano J.R. (2024), The impact of Artificial Intelligence in the financial sector: opportunities and challenges, *International Journal of Business & Management Studies*, 05, n.10, pp.33-42

Bessen J. (2017), Automation and jobs: when technology boost employment, *Boston University School of Law, Law and Economics Research Paper*, n.17-09, Boston, Boston University

Camassa S., Ferri V., Perego S., Porcelli R. (2024), Gli effetti dell'AI sulle competenze: come decifrarli con l'analisi degli annunci di lavoro, *AgendaDigitale.ue*, 3 maggio

Carbonero F., Ernst E., Weber E. (2018), *Robots worldwide: the impact of automation on employment and trade*, IAB-Discussion Paper n.2020/07, Nürnberg, Institute for Employment Research of the Federal Employment Agency

Chiacchio F., Petropoulos G., Pichler D. (2018), *The impact of industrial robots on EU employment and wages. A local labour market approach*, Bruegel Working Paper n.2, Brussels-Bruegel

Dalla Zuanna A., Dottori D., Gentili E., Lattanzio S. (2024), *An assessment of occupational exposure to artificial intelligence in Italy,* Bank of Italy Occasional Paper n.878, Roma, Banca d'Italia

De Backer K., DeStefano T. (2021), Robotics and the global organisation of production, in von Braun J., Reichberg G.M., Archer M. (eds.), *Robotics, AI, and Humanity: Science, Ethics, and Policy*, Cham, Springer International Publishing, pp.71-84

Duygun M., Hashem S.Q., Tanda A. (2021), Editorial: Financial intermediation versus disintermediation: Opportunities and challenges in the FinTech era, *Frontiers in Artificial Intelligence*, 3, Article 629105

Ernst E., Merola R., Samaan D. (2019), Economics of artificial intelligence: implications for the future of work, *IZA Journal of Labor Policy*, 9, n.1, pp.1-35

Farazi N.Z.R. (2024), Evaluating the impact of AI and blockchain on credit risk mitigation: A predictive analytic approach using machine learning. *International Journal of Science and Research Archive*, 13, n.1, pp.575-582

Felten E., Raj M., Seamans R. (2021), Occupational, industry, and geographic exposure to artificial intelligence: a novel dataset and its potential uses, *Strategic Management Journal*, 42, n.12, pp.2195-2217

Ferri V., Porcelli R., Fenoaltea E.M. (2024), *Lavoro e Intelligenza Artificiale in Italia: tra opportunità e rischio di sostituzione*, Inapp Working Paper n.125, Roma, Inapp

Frey C.B., Osborne M.A. (2017), The future of employment. How susceptible are jobs to computerisation?, *Technological forecasting and social change*, 114, pp.254-280

Graetz G., Michaels G. (2015), *Robots at work*, IZA Discussion Paper n.8938, Bonn, IZA

Guarascio D., Reljic J., Stöllinger R. (2023), *Artificial intelligence and employment: a look into the crystal ball*, GLO Discussion Paper n.1333, Essen, Global Labor Organization

Huang M.H., Rust R.T. (2018), Artificial intelligence in service, *Journal of Service Research*, 21, n.2, pp.155-172

Jagdale R., Deshmukh M. (2025), *Natural Language Processing in Finance: Techniques, Applications, and Future Directions*, in Chen H. (ed.), *Machine Learning and Modeling Techniques in Financial Data Science*, Hershey (PA), IGI Global, pp.411-434

Jiang W., Tang Y., Xiao R.J., Yao V. (2021), Surviving the fintech disruption, NBER Working Paper n.28668, Cambridge, MA, NBER

Kaneria G. (2022), Artificial Intelligence in Banking Industry, in Balamurugan S., Pathak S., Jain A., Gupta S., Sharma S., Duggal S. (eds.), *Impact of Artificial Intelligence on Organizational Transformation*, Hoboken (NJ), John Wiley & Sons, pp.327-348

Kour M., Kour R. (2024), AI and Influencer Marketing: Redefining the Future of Social Media Marketing in Industry 6.0, In *Advanced Businesses in Industry 6.0*, Hershey (PA), IGI Global, pp.87-103

Mazzarella R., Mallardi F., Porcelli R. (2017), Atlante lavoro, *Sinappsi*, n.2/3, pp.7-26

Paul L.R., Sadath L., Madana A. (2021), Artificial Intelligence in Predictive Analysis of Insurance and Banking, in Bhargava C., Sharma P.K. (eds.), *Artificial Intelligence*, Boca Raton (FL), CRC Press, pp. 31-54

Pizzinelli C., Panton A.J., Mendes Tavares M., Cazzaniga M., Li L. (2023), *Labor market exposure to AI: cross-country differences and distributional implications*, Working Paper n.2023/216, Washington, International Monetary Fund

Priya B., Sharma V. (2023), Exploring Users'Adoption Intentions of Intelligent Virtual Assistants in Financial Services. An Anthropomorphic Perspectives and Socio-Psychological Perspectives, *Computers in Human Behavior,* 148, Article 107912

Rahman M., Ming T.H., Baigh T.A., Sarker M. (2021), Adoption of Artificial Intelligence in Banking Services. An Empirical Analysis. *International Journal of Emerging Markets* <DOI:10.1108/IJOEM-06-2020-0724>

Singh S. (2019), Measuring E-Service Quality and Customer Satisfaction with Internet Banking in India, *Theoretical Economics Letters*, 9, pp.308-326

Skandali D., Magoutas A., Tsourvakas G. (2023), Artificial intelligent applications in enabled banking services: the next frontier of customer engagement in the era of ChatGPT, *Theoretical Economics Letters*, 13, n.5, pp.1203-1223

Thorp H.H. (2023), ChatGPT is fun, but not an author, *Science*, 379, n.6630, pp.313-313

Vivarelli M. (2014), Innovation, employment and skills in advanced and developing countries: a survey of economic literature, *Journal of Economic Issues*, 48, n.1, pp.123-154

Andrea Battistoni
a.battistoni@inapp.gov.it
Architetto e dottore di ricerca in Geostoria e Geografia. Portavoce del Presidente Inapp, esperto di comunicazione istituzionale e innovazione sociale. Ha ricoperto ruoli strategici presso Presidenza del Consiglio, CNEL e Ministero del Lavoro. Tra le sue pubblicazioni più importanti *La rigenerazione urbana come politica pubblica per lo sviluppo dei territori* (2024); *Politiche pubbliche e infrastrutture sociali: una visione integrata per il benessere* (2024); *Verso una nuova normalità di partecipazione* (2025)

Valentina Ferri
v.ferri@inapp.gov.it
Dottore di ricerca in Economia della popolazione e dello sviluppo, è ricercatrice presso l'Inapp, dove si occupa di formazione continua, politiche del lavoro e transizione istruzione-lavoro. Coordina convenzioni riguardanti le trasformazioni strutturali sul mercato del lavoro e sulla formazione. Fra le pubblicazioni recenti: *L'impatto dell'intelligenza artificiale sulle competenze e sulla formazione professionale* (2024); *Gli effetti dell'AI sulle competenze: come decifrarli con l'analisi degli annunci di lavoro* (2024); *L'impatto del Fondo nuove competenze sulla Formazione continua. Evidenze dai dati INDACO-Imprese 2022* (2024).

Tecnologie IA, imprese e domanda di lavoro

Irene Brunetti
INAPP

Andrea Ricci
INAPP

L'articolo indaga l'esistenza e la natura della relazione tra gli investimenti nelle tecnologie di Intelligenza artificiale (IA) e la domanda di lavoro delle imprese italiane. In tale prospettiva si utilizzano le informazioni contenute nella componente longitudinale della Rilevazione Imprese e lavoro (RIL). Le elaborazioni econometriche mostrano che l'adozione di tecnologie IA non influenza in modo significativo la quota di nuove assunzioni (domanda effettiva), mentre si associa a un debole incremento della quota di posti di lavoro vacanti (domanda potenziale). Queste evidenze tengono conto delle traiettorie di digitalizzazione che caratterizzano le singole imprese e riflettono soprattutto il comportamento delle realtà produttive di medio-grandi dimensioni. Nel complesso, i risultati supportano l'idea che la diffusione di ecosistemi di IA – sebbene sia ancora molto limitata – tende ad accelerare la polarizzazione del mercato del lavoro e il dualismo competitivo del sistema imprenditoriale.

This paper investigates the existence and nature of the relationship between investments in Artificial intelligence (AI) technologies and the labour demand of Italian firms. To this end, the study draws on data from the longitudinal component of the Rilevazione Imprese e lavoro (RIL) survey. Our econometric estimations indicate that the adoption of AI technologies does not significantly affect the share of new hires (effective demand), while it is associated with a slight increase in the share of job vacancies (potential demand). These findings account for the individual digitalisation trajectories of firms and primarily reflect the behaviour of medium- and large-sized enterprises. Overall, our findings support the notion that the diffusion of AI ecosystems – although still limited in scope – tends to accelerate labour market polarisation and the competitive dualism within the Italian productive system.

DOI: 10.53223/Sinappsi_2025-02-13

Citazione
Brunetti I., Ricci A. (2025), Tecnologie IA, imprese e domanda di lavoro, *Sinappsi*, 15, n.2, pp.192-207

Parole chiave
Domanda di lavoro
Intelligenza artificiale
Imprese

Keywords
Labour demand
Artificial intelligence
Firms

Introduzione

Il dibattito accademico e istituzionale negli ultimi anni si è concentrato sempre più frequentemente sulle tecnologie di Intelligenza artificiale (IA) e sulle implicazioni che esse hanno per le prospettive del mercato del lavoro e della

dinamica economico-sociale (Brynjolfsson *et al.* 2018; Acemoglu e Restrepo 2018; Acemoglu *et al.* 2022; Agrawal *et al.* 2023; Schwab 2025).

L'Intelligenza artificiale è oggi caratterizzata da una crescente digitalizzazione, interconnettività e automazione dei processi economici e agisce come una tecnologia di uso generale che sta ridefinendo il modello di business e le strategie competitive delle aziende. Le tecnologie IA sono associate a sistemi e macchine che tendono a eseguire/replicare alcune funzioni cognitive tipiche degli esseri umani (ad esempio, apprendere, comprendere, ragionare e interagire), ovvero elaborare previsioni e prendere decisioni sulla base di obiettivi e finalità di varia natura (OECD 2019).

In questa prospettiva alcuni studi sottolineano come l'adozione di sistemi di IA possa generare processi di innovazione cumulativi e una maggiore efficienza nello svolgimento di attività a elevato contenuto cognitivo, con un conseguente incremento della produttività e della domanda per alcuni profili professionali (Cockburn *et al.* 2019; Brynjolfsson *et al*. 2025; Brynjolfsson *et al.* 2021; Bianchini *et al.* 2022; Noy e Zhang 2023; Calvino e Fontanelli 2024; Brunetti *et al*. 2025).

La diffusione delle tecnologie IA, d'altra parte, potrebbe risolversi nella sostituzione di un numero crescente di lavoratori, nella misura in cui la loro applicazione permette di svolgere mansioni e compiti sempre più complessi, rimodellando strutturalmente i processi produttivi e organizzativi nella direzione di un minore impiego di lavoro (Acemoglu e Restrepo 2019; Brynjolfsson e McAfee 2014; Agrawal *et al.* 2022). In tale circostanza si può verificare una riduzione significativa della quota del lavoro sul reddito nazionale e un aumento delle disuguaglianze, come pure tensioni sulle dinamiche di inclusione sociale (Acemoglu e Restrepo 2018). L'adozione e l'integrazione dell'IA nei processi produttivi chiama in causa, inoltre, il tema fondamentale della riqualificazione e rimodulazione delle competenze della forza lavoro, la governance dei dati per la protezione delle informazioni sensibili (Bostrom e Yudkowsky 2014), nonché sfide fondamentali in ambito etico, geopolitico e dei processi democratici.

Nonostante la rilevanza politico-istituzionale e l'attenzione della letteratura scientifica sul tema, va sottolineato che conosciamo ancora poco dei meccanismi microeconomici che sono alla base degli investimenti in IA del mondo imprenditoriale, ovvero delle implicazioni di queste tecnologie per la domanda di lavoro e per la creazione e distruzione dei fabbisogni professionali (Alekseeva *et al.* 2021; Babina *et al.* 2024). Ciò non sorprende se consideriamo che parliamo di una dinamica tecnologica relativamente recente e di fattori e comportamenti estremamente complessi, il tutto in presenza di una limitata disponibilità dei microdati sulle caratteristiche delle imprese – e dei lavoratori ivi occupati.

Indubbiamente la propensione a investire in IA riflette una molteplicità di elementi, che non si esauriscono nella presenza di infrastrutture tecnologiche, computazionali e organizzative a livello aziendale, così come nell'esercizio di pratiche manageriali che tendono a utilizzare i dati per ottimizzare il modello di business. L'adozione delle nuove tecnologie riflette altri aspetti che riguardano il profilo di conoscenze e competenze degli occupati, la specializzazione produttiva, l'assetto delle relazioni industriali e le norme implicite che regolano i rapporti di lavoro[1] (Agrawal *et al.* 2019; Igna e Venturini 2023; Babina *et al.* 2024).

In questo contesto, le traiettorie tecnologiche dell'automazione e della digitalizzazione, congiuntamente alle pratiche manageriali e all'organizzazione delle risorse umane, interagiscono tra loro nel condizionare le modalità attraverso cui i sistemi IA influenzano il livello e la natura della domanda di lavoro. Secondo Agrawal *et al.* (2018), ad esempio, poiché l'IA può aumentare la produttività dei lavoratori, specialmente nei settori ad alta intensità di conoscenza, gli investimenti in IA stimolerebbero la domanda di lavoro qualificato capace di gestire e ottimizzare le nuove tecnologie.

A fronte di questa complessità, emerge chiaramente la necessità di collocare le analisi

1 Alcune ricerche hanno illustrato come l'IA tenda a manifestare i suoi effetti in due aree principali: l'automazione delle attività ripetitive e la valorizzazione delle competenze umane attraverso strumenti avanzati. Software e robot, ad esempio, hanno sostituito i lavoratori poco o mediamente qualificati (Autor *et al.* 2003; Acemoglu e Restrepo 2020), mentre l'arrivo di alcuni macchinari ha portato, come nel settore dell'elettricità, a una completa riorganizzazione dei processi di produzione all'interno delle aziende (Juhász *et al.* 2024).

del rapporto tra investimenti in IA e domanda di lavoro nel contesto di un ordito analitico, di una strategia empirica e di una base statistico-informativa adeguata a cogliere le molteplicità e complessità dei fattori che influenzano gli effetti delle nuove tecnologie nei mercati interni del lavoro.

In quest'ottica, le analisi presentate nel nostro studio si caratterizzano per l'assunzione di una prospettiva microeconomica e per l'attenzione rivolta alle scelte di investimento con dati di impresa, un ambito empirico ancora poco esplorato dalla letteratura sull'argomento (Calvino e Fontanelli 2023; Babina *et al.* 2024), spesso connotata da studi su dati aggregati e/o da evidenze indirette riguardanti l'esposizione al rischio di automazione di lavoratori e professioni (Albanesi *et al.* 2025; Guarascio e Reljic 2025; Bonfiglioli *et al.* 2025; Dalla Zuanna *et al*. 2024). La nostra ricerca si focalizza in particolare sulle implicazioni delle scelte di investimento in tecnologie IA sulle nuove assunzioni e sull'intensità della ricerca di nuove figure professionali. I risultati ottenuti a partire dai dati della V e VI Rilevazione Imprese e lavoro (RIL-2018 e RIL-2022) suggeriscono che l'investimento, in sistemi IA non influenza in modo significativo la quota di assunzioni (domanda effettiva), mentre si associa a un debole incremento della quota di posti vacanti (domanda potenziale). Tali evidenze riflettono soprattutto il comportamento delle imprese di medio-grandi dimensioni. Nonostante i dati utilizzati non permettano di identificare le caratteristiche dei neoassunti e il profilo dei fabbisogni professionali, i nostri risultati supportano l'idea che la diffusione delle tecnologie IA – sebbene ancora molto limitata all'interno del tessuto produttivo italiano – tenda ad accelerare la polarizzazione del mercato del lavoro e il dualismo competitivo del sistema imprenditoriale. Tale interpretazione è peraltro coerente con studi recenti nei quali si dimostra come nel nostro Paese le aziende che utilizzano IA offrano mediamente salari più elevati e tendano ad assumere lavoratori in occupazioni ad alto contenuto cognitivo (Ciani *et al*. 2025).

L'articolo è strutturato come segue: il paragrafo 1 presenta, alla luce della letteratura esistente, una discussione preliminare; il 2 introduce i dati e le statistiche descrittive; il paragrafo 3 descrive il modello di regressione utilizzato; nel paragrafo 4 vengono discussi i risultati principali, mentre l'ultimo conclude.

1. Una discussione preliminare

La cosiddetta quarta rivoluzione industriale sembra destinata ad esercitare un impatto sostanziale sull'evoluzione del tessuto produttivo e del mercato del lavoro, con implicazioni evidenti sulla dinamica dell'occupazione e sulla domanda di competenze espresse dal sistema delle imprese (Bessen 2017; Acemoglu e Restrepo 2020; OECD 2023; Drydakis 2025). La letteratura economica che analizza empiricamente la relazione tra Intelligenza artificiale e mercato del lavoro è infatti in rapida crescita e riguarda diversi ambiti. Vi sono studi che analizzano gli effetti delle nuove tecnologie sull'occupazione giungendo, tuttavia, a risultati contrastanti: l'IA può infatti agire sia come complemento del lavoro umano, sia come suo sostituto (Autor e Salomons 2018). Graetz e Michaels (2015), ad esempio, non rilevano effetti significativi dell'utilizzo dei robot sull'occupazione, mentre altri evidenziano che un loro impiego può portare a una perdita di posti di lavoro, in particolare per i lavoratori a basse competenze che svolgono professioni routinarie, e a una riduzione dei salari (Acemoglu e Restrepo 2020). Altre ricerche si sono concentrate sul concetto di 'esposizione' delle occupazioni all'Intelligenza artificiale. Le definizioni proposte di esposizione considerano in che misura le applicazioni dell'IA si sovrappongano alle capacità umane necessarie per svolgere una determinata occupazione (come nell'Indice AIOE di Felten *et al.* 2021 e 2023) oppure quanto l'IA possa accelerare significativamente l'esecuzione dei compiti previsti in ciascun lavoro (Webb 2019; Eloundou *et al.* 2023). Tale definizione, tuttavia, non consente di distinguere tra la possibilità che l'IA funzioni come sostituto o come complemento del lavoro umano nei compiti chiave, o che possa eventualmente sostituire del tutto un'occupazione. Pizzinelli *et al.* (2023) propongono un'estensione dell'Indicatore di Felten *et al*. (2021), tenendo conto del potenziale dell'Intelligenza artificiale sia come complemento, sia come sostituto del lavoro umano.

Un altro filone della letteratura riguarda le implicazioni per la forza lavoro nel suo complesso. Da un lato, l'IA può migliorare la qualità del lavoro, riducendo l'incidenza di compiti usuranti e aprendo nuove possibilità professionali.

Dall'altro, il rischio di polarizzazione del mercato del lavoro è reale: i lavoratori con titolo di istruzione e competenze digitali elevate tendono a beneficiare dell'automazione, mentre coloro che non possiedono tali competenze possono sperimentare disoccupazione, sottoccupazione o un deterioramento delle condizioni lavorative (Brynjolfsson e McAfee 2014; OECD 2023; Draca *et al.* 2024).

Dal punto di vista delle imprese, le evidenze empiriche sulla diffusione degli investimenti in IA e il loro impatto sulle scelte di assunzione sono ancora limitate a causa della relativa novità dei progressi tecnologici in questo campo, e della scarsa disponibilità di dati a livello micro (Calvino e Fontanelli 2023). La ricerca sui dati a livello di impresa si basa principalmente su quattro diverse fonti: indagini ufficiali su Information and Communication Technology condotte dagli uffici nazionali di statistica; dati online sui posti di lavoro vacanti; registri di proprietà intellettuale o attività brevettuale delle imprese; infine, esperimenti sul campo e/o studi di casi. La nostra analisi fa parte dalla prima tipologia[2]. Zolas *et al.* (2020) suggeriscono che, nel 2018, solo il 6,6% delle imprese statunitensi faceva uso delle tecnologie di IA. Evidenze analoghe emergono anche per i Paesi manifatturieri europei: Czarnitzki *et al.* (2023) mostrano, ad esempio, che nel 2018 l'uso delle tecnologie IA nelle imprese tedesche era di circa il 5,8%, concentrato nelle imprese più grandi e in alcuni settori di attività. Allo stesso modo, Cho *et al.* (2022) evidenziano che il 9% delle aziende coreane utilizza la tecnologia dell'Intelligenza artificiale e che la sua adozione è legata ad altre infrastrutture e capacità digitali.

Con riferimento all'Italia, l'Istat (2021) mostra che circa il 6,2% delle imprese con più di 10 addetti ha utilizzato sistemi di IA per almeno una delle sette finalità proposte (8% la media UE), quota che sale al 15,4% tra le imprese dei settori ICT e che raggiunge incidenze più elevate nelle telecomunicazioni (18,1%) e nelle tecnologie dell'informazione (15,7%). Sempre per l'Italia, il Rapporto Inapp (Inapp 2024) evidenzia che l'incidenza media delle imprese che hanno investito effettivamente in tecnologie di IA nel periodo 2019-2021 è molto più limitata e, soprattutto, distribuita in modo disomogeneo rispetto a le dimensioni delle imprese, i settori di attività e le macroregioni. L'adozione delle tecnologie di Intelligenza artificiale varia, inoltre, tra le aziende anche a causa di altri fattori come le caratteristiche gestionali, le relazioni industriali, la concorrenza dei lavoratori e le dotazioni di capitale umano. Ad esempio, vi è evidenza che la possibilità che l'adozione pregressa di varie tecnologie digitali (Big Data, Internet of things, Robotica) generi un successivo investimento in sistemi di IA – secondo una logica di *path depencency* tecnologica – è fortemente condizionata (negativamente) dalla presenza di management dinastico – a sua volta riflesso della grande prevalenza di imprese piccole in dimensioni, e familiari nella proprietà. Questo risultato supporta l'idea che, in un sistema produttivo come quello italiano, l'analisi delle implicazioni degli investimenti in IA su diversi *outcomes* del mercato dal lavoro e, specificamente, sull'evoluzione della domanda di lavoro, debba essere declinata in un ecosistema digitale che tenga conto esplicitamente delle caratteristiche manageriali, organizzative e produttive.

Rispetto al legame tra gli investimenti in tecnologie digitali, tra cui IA, e la domanda di lavoro, la teoria della crescita endogena (i.e., Kogan *et al.* 2017) spiega che le imprese con innovazioni di successo tendono a crescere in termini di dimensioni, il che porta naturalmente a maggiori investimenti a livello aziendale e viceversa. Inoltre, l'analisi dell'impatto dell'innovazione e degli investimenti sulle decisioni di assunzione non è chiara. Diversi studi – tra cui Aghion *et al.* (2020), Acemoglu *et al.* (2023b), Bratta *et al.* (2023) – evidenziano un impatto positivo degli investimenti legati all'automazione sull'occupazione nelle imprese che investono, sebbene con effetti differenziati a seconda delle competenze e delle mansioni dei lavoratori coinvolti. Acemoglu e Restrepo (2018) mostrano come l'occupazione potrebbe aumentare a seguito dell'introduzione di nuove tecnologie. Anche Babina *et al.* (2024) documentano che le aziende che investono in IA sperimentano aumenti nella loro occupazione complessiva. Se guardiamo a un gruppo particolare di lavoratori, i giovani, Battisti *et al.* (2024) suggeriscono che essere stati esposti, ovvero impiegati in un'impresa che ha scelto di investire in nuove tecnologie, all'automazione è positivamente

2 Le variabili che descrivono l'utilizzo delle tecnologie di Intelligenza artificiale sono generalmente binarie: alle aziende viene chiesto se utilizzano o meno sistemi di IA – generalmente senza informazioni rilevanti sui tempi di adozione.

associato alla probabilità di rimanere occupati, migliorando quindi le prospettive occupazionali di questo specifico segmento della forza lavoro. Esistono, tuttavia, altre ricerche che giungono a conclusioni opposte, suggerendo che gli investimenti in automazione possano danneggiare i livelli occupazionali all'interno delle aziende (i.e., Bessen *et al.* 2023; Bonfiglioli *et al.* 2023). Acemoglu *et al.* (2023a) non riscontrano cambiamenti significativi nei tassi di crescita dell'occupazione a seguito degli investimenti in nuove tecnologie digitali concludendo che il fattore determinante per l'occupazione osservata sia soprattutto l'auto-selezione delle imprese verso l'investimento. Con riferimento all'Italia, i risultati di Caselli *et al.* (2024) indicano che l'investimento in tecnologie informatiche digitali influisce positivamente sull'occupazione, favorendo una forza lavoro qualificata e competente in IT, al contrario, l'investimento in tecnologie operative non altera significativamente l'occupazione totale. La mancanza di consenso su questo tema giustifica la necessità di nuove analisi. Il presente articolo mira a colmare tale divario, analizzando la risposta delle aziende in termini di assunzioni e ricerca di nuovo personale in risposta agli investimenti in nuove tecnologie.

2. Dati e statistiche descrittive

I dati

L'analisi empirica si sviluppa a partire dai dati della componente longitudinale della Rilevazione su Imprese e lavoro (RIL) condotta periodicamente dall'Inapp su un campione rappresentativo di società di capitali e di persone operanti nel settore privato extra-agricolo[3]. L'indagine RIL contiene un ricco insieme di informazioni relative agli assetti manageriali e di governance societaria, all'organizzazione del lavoro e alle politiche del personale, alla composizione della forza lavoro occupata, alle relazioni industriali, alle scelte di investimento e ai comportamenti competitivi.

La ricerca si basa sulle ultime due edizioni dell'indagine RIL. Da RIL-2018 acquisiamo i dati sul processo di digitalizzazione avviato dalle imprese nel triennio 2015-2017, mentre da RIL-2022 le informazioni sull'adozione di tecnologie di Intelligenza artificiale nel periodo 2019-2021. Nella fattispecie il questionario dell'indagine RIL-2022 include la seguente domanda: "*Nel periodo 2019-2021 l'impresa ha effettuato investimenti in beni materiali o immateriali o acquistato servizi relativi alle seguenti aree tecnologiche da utilizzare per le proprie attività*". L'impresa può scegliere tra le seguenti opzioni: (*i*) Soluzioni di Internet delle cose e/o Realtà aumentata e virtuale (IoT); (*ii*) Stampanti 3D; (*iii*) Robotica; (*iv*) Cloud computing e analisi dei Big Data; (*v*) Applicazioni web o dispositivi per la vendita online; (*vi*) Cybersecurity e aggiornamenti tecnologici dei dispositivi esistenti; (*vii*) Intelligenza artificiale; (*viii*) Altro. Queste categorie rappresentano le principali tecnologie abilitanti del pacchetto Industria 4.0 fondamentale per la trasformazione digitale delle imprese.

Per quanto concerne i flussi di lavoratori in entrata e in uscita, nonché la ricerca di personale, un elemento importante è capire come misurare, a livello microeconomico, la domanda di lavoro: se riferirsi alla variazione effettiva della quota di occupati (assunzioni e cessazioni del numero di lavoratori) oppure alla variazione della quota di posti vacanti. Il tasso dei posti vacanti è uno dei principali indicatori economici sull'evoluzione del mercato del lavoro nella misura in cui può avere un valore predittivo rispetto alle oscillazioni del tasso di occupazione. Nella nostra analisi facciamo riferimento a due variabili principali: la quota di nuovi lavoratori assunti con contratto dipendente calcolata sul totale della forza lavoro impiegata, e la quota di personale che l'impresa sta attualmente ricercando (occupazione 'potenziale') per assunzioni

3 La popolazione cui si riferisce l'indagine RIL è rappresentata dalle imprese attive in tutti i settori privati non agricoli senza alcuna limitazione sulla dimensione di impresa. Le imprese intervistate vengono estratte casualmente utilizzando la banca dati ASIA (Archivio statistico delle imprese attive) fornita dall'Istat, secondo una strategia di campionamento volta a costruire un campione statisticamente rappresentativo delle imprese attive in Italia. La stratificazione è ottenuta dalla concatenazione delle variabili: classe dimensionale di impresa (in termini di media annua di addetti), regione della sede legale dell'impresa e settore di attività economica (secondo una aggregazione della classificazione Ateco 2007). La metodologia di calcolo dei pesi di riporto all'universo e il loro conseguente utilizzo assicura la validità esterna dei risultati ottenuti. La rilevazione prevede una quota longitudinale di circa il 35%, vale a dire un sottocampione di imprese coinvolte nell'indagine precedente.

con contratto dipendente calcolata rispetto alla media dell'occupazione effettiva (e, in alternativa seguendo la definizione data da Eurostat, identificata dalla somma tra i dipendenti e i posti vacanti misurati nell'anno corrente).

Queste variabili catturano due diverse misure di domanda di lavoro. La quota delle assunzioni si riferisce all'annualità 2021, mentre la quota di posti vacanti è misurata al momento dell'intervista – ovvero tra febbraio e dicembre 2022. La quota di assunzioni è, inoltre, una grandezza effettiva, mentre la quota di posti vacanti è una variabile percettiva che include le aspettative degli imprenditori e le strategie di impresa per il prossimo futuro. Tale distinzione è fondamentale quando si analizzano i processi di cambiamento strutturale come la digitalizzazione e gli investimenti in sistemi di IA. Si tratta di fenomeni che tipicamente tendono ad accompagnarsi a processi di apprendimento che, dal punto di vista statistico, hanno una debole memoria storica. In altre parole, se la finalità è prevedere le implicazioni occupazionali e sociali della transizione IA, l'evoluzione dei posti vacanti può essere un indicatore più informativo rispetto ai flussi dei rapporti di lavoro.

La ricchezza delle informazioni contenute in RIL permette, quindi, di esaminare la relazione tra l'investimento in IA e la domanda di lavoro controllando rispetto a numerosi fattori di potenziale confondimento: l'introduzione di tecnologie digitali – che possono sovrapporsi ai processi di efficientamento energetico –, i processi di riorganizzazione conseguenti lo shock Covid e altri elementi microeconomici che possono influenzare le scelte di ampliamento e/o riduzione della base occupazionale (i.e., presenza sui mercati internazionali, blocco dei licenziamenti introdotto a seguito della pandemia, competenze manageriali ecc.). Indubbiamente le caratteristiche produttive, manageriali e competitive possono incidere sulle politiche del personale indipendentemente dalla natura e dall'entità dell'investimento in tecnologie della transizione. Al fine di esaminare mercati interni del lavoro con un livello minimo di struttura organizzativa, si selezionano solo le imprese con almeno 10 dipendenti. Una volta imposto tale criterio ed eliminate le osservazioni con valori mancanti, si ottiene un campione longitudinale di oltre 7.300 imprese.

Le statistiche descrittive

La tabella 1 presenta le statistiche descrittive relative alla domanda di lavoro, alla diffusione degli investimenti in Intelligenza artificiale e nelle tecnologie abilitanti, nonché la media e la deviazione standard relative alle altre caratteristiche dell'impresa e della forza lavoro occupata per il campione longitudinale. Con riferimento alla domanda di lavoro, la quota di neoassunti nel 2021 è pari al 23% dei lavoratori occupati, a fronte di un tasso di posti vacanti che nell'anno seguente si attesta tra il 4,2% e 3,4%, a seconda che il numero di posti vacanti sia calcolato rapportandolo alla media dei lavoratori occupati o al totale delle posizioni lavorative (occupate e vacanti). L'evoluzione positiva della domanda di lavoro si evince dalle statistiche sulla quota di neoassunti (17,7%) e di posti vacanti (2,0% e 1,7%) che si riferiscono alle annualità 2018 e 2019, rispettivamente. Si tratta di un'evidenza in linea con i dati dell'Istat secondo cui il tasso medio di posti vacanti è progressivamente aumentato negli ultimi anni passando da un valore medio di circa 1% nel periodo pre-Covid-19 ad un 2% nel biennio 2023-2023 (Istat - Indagine Vela). La tabella 1 evidenzia che solo l'1,4% delle imprese ha investito in tecnologie di Intelligenza artificiale tra il 2019 e il 2021. Questo dato riflette la composizione del tessuto produttivo italiano, dominato da piccole e micro imprese poco inclini all'adozione di nuove tecnologie, a fronte di una minoranza di grandi aziende che invece ha iniziato a introdurle.

Il processo di digitalizzazione che esaminiamo è fortemente connesso al Piano Industria 4.0 del periodo 2015-2017 e si manifesta in una diffusione limitata delle tecnologie abilitanti: Big Data e Cloud computing (7%), Internet delle cose e Realtà aumentata (11%), Robotica e Smart manufacturing (6%). Questi dati confermano elementi già noti della dinamica tecnologica nelle imprese italiane, dinamica caratterizzata da una debole propensione innovativa e da strategie di tipo *single technology* piuttosto che *multiple-technologies* – basate, cioè, su investimenti simultanei in tecnologie complementari (Cirillo *et al.* 2023). Le evidenze descrittive suggeriscono quindi di considerare il paradigma IA come 'esito' di una traiettoria tecnologica (*path dependence*) declinandola nell'ambito di un complesso ecosistema caratterizzato da debolezze strutturali nel processo di introduzione delle tecnologie digitali.

Tabella 1. Statistiche descrittive

	Media	Dev std	Min	Max
		Domanda di lavoro		
Quota assunti $_{2021}$	23%	0,41	0	1
Quota posti vacanti $_{2022}$	4%	0,09	0	1
Quota vacanti (Eurostat) $_{2022}$	3%	0,06	0	1
Quota assunti $_{2017}$	18%	0,32	0	1
Quota posti vacanti $_{2018}$	2%	0,05	0	1
Quota vacanti (Eurostat) $_{2018}$	2%	0,04		1
		Tecnologie digitali		
AI $_{2019\text{-}21}$	1%	0,12	0	1
Big Data (Cloud computing)$_{2015\text{-}17}$	7%	0,26	0	1
Internet delle cose (Realtà aumentata)$_{2015\text{-}17}$	11%	0,32	0	1
Robotica$_{2015\text{-}17}$	6%	0,24	0	1
		Caratteristiche management		
Istruzione terziaria	28%	0,45	0	1
Istruzione secondaria superiore	51%	0,50	0	1
Età (in anni)	57	11,47	20	99
Manager donna	13%	0,34	0	1
Proprietà familiare	85%	0,36	0	1
		Caratteristiche occupati		
Quota dirigenti	4%	0,10	0	1
Quota impiegati	36%	0,31	0	1
Quota operai	59%	0,33	0	1
Quota donne	34%	0,25	0	1
Quota laureati	14%	0,21	0	1
Quota diplomati	49%	0,29	0	1
Quota licenza media-elementare	38%	0,32	0	1
		Caratteristiche imprese		
Incentivi fiscali per investimenti	30%	0,46	0	1
Blocco licenziamenti-Covid*	3%	0,18	0	1
Mercati internazionali (export)	37%	0,48	0	1
Età impresa (espressa in anni)	29	17,09	1	192
Ln (fatturato per dipendente)	11,83	1,26	6,11	15,22
Lavoro da casa- Covid*	23%	0,42	0	1
Contrattazione II livello	10%	0,30	0	1
Industria	47%	0,50	0	1
1-9 dip	-	-	-	-
10-49 dip	84%	0,36	0	1
50-249 dip	13%	0,34	0	1
>250 dip	3%	0,16	0	1
N. Osservazioni	7.343			

Nota: applicazione pesi longitudinali.

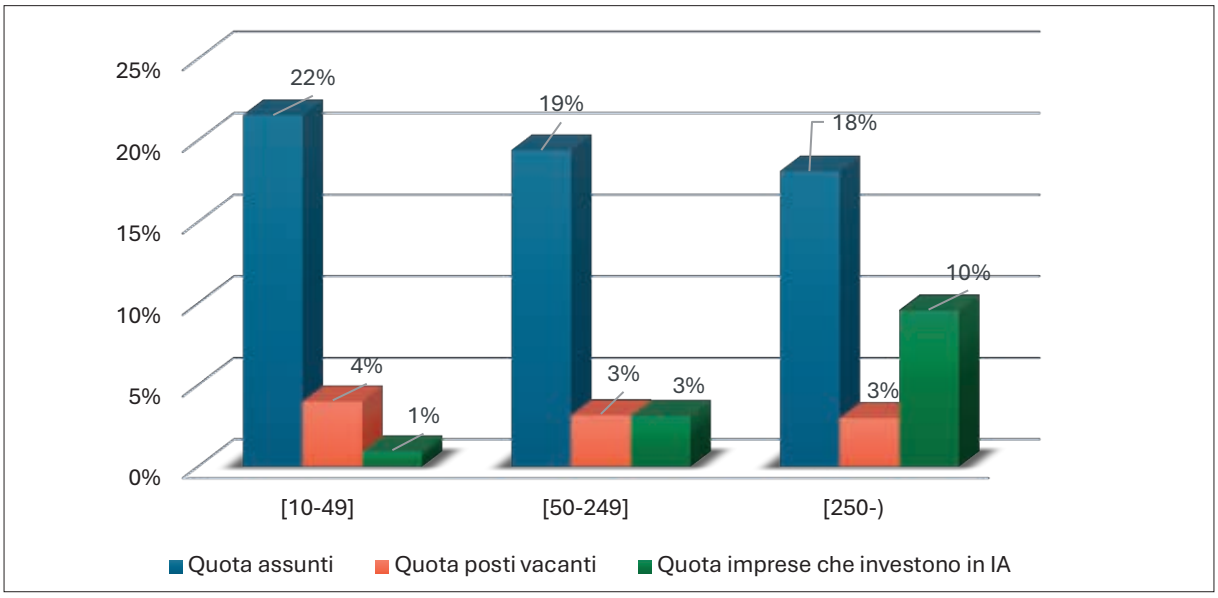
Fonte: elaborazioni degli Autori su dati RIL 2018 e 2022

La tabella 1 riporta, inoltre, le statistiche per le altre variabili utilizzate nell'analisi multivariata: il profilo manageriale e di *corporate governance*, la composizione della forza lavoro occupata e le caratteristiche produttive. Le evidenze sono coerenti con quanto emerge in studi precedenti circa il sistema imprenditoriale italiano: l'incidenza media di manager laureati e manager donne è limitata (28% e 13%, rispettivamente), mentre è elevata la loro età media (57 anni) e la quota di proprietà familiare (85%). In merito alla composizione della forza lavoro, si rileva che la quota di lavoratori con un titolo di studio universitario (14%) è largamente inferiore alla proporzione dei dipendenti con istruzione secondaria superiore (49%), a conferma di come il basso livello medio di istruzione sia un aspetto comune degli assetti manageriali e della forza lavoro occupata. Le statistiche relative alle caratteristiche di impresa, infine, confermano alcuni aspetti noti del sistema produttivo italiano. Le aziende presenti nel campione RIL sono in media di piccole dimensioni, poco esposte al commercio internazionale (37%) e scarsamente inclini ad applicare la contrattazione integrativa (10%). Lo shock Covid-19 ha avuto conseguenze significative per la quota di imprese che ha dovuto affrontare la riorganizzazione del lavoro (23%).

È ben noto che il tessuto imprenditoriale italiano è caratterizzato da elevata eterogeneità connessa alla prevalenza delle piccole e medie imprese (PMI) gestite da management dinastico ed espressione di proprietà familiare. Questi aspetti condizionano una molteplicità di fattori (strategie competitive, organizzazione dei mercati interni, norme sociali implicite) che direttamente e indirettamente influenzano la propensione innovativa e la penetrazione degli ecosistemi digitali, ovvero la dinamica della domanda di lavoro e di nuove competenze.

La figura 1 illustra la distribuzione della domanda di lavoro e degli investimenti in IA in rapporto alla dimensione di impresa. In linea con le attese, gli investimenti in IA sono concentrati nelle imprese con oltre 250 dipendenti (10%) mentre sono marginali nelle piccole aziende (1%) che, come già detto, identificano la componente preponderante del nostro tessuto produttivo. L'intensità della ricerca di nuovo personale, invece, è relativamente più elevata tra le piccole rispetto alle medio-grandi imprese (4% vs 3%). La figura 1 può, d'altra parte, essere condizionata dall'avvio della ripresa economica post-pandemica e dall'introduzione di diverse misure di incentivazione fiscale per le assunzioni (Inapp 2024). Il quadro descrittivo viene arricchito

Figura 1. Quota di assunzioni, posti vacanti e di imprese che investono in IA per classe dimensionale



Nota: applicazione pesi longitudinali.

Fonte: elaborazioni degli Autori sul campione longitudinale RIL-2018 e RIL-2022

illustrando l'evoluzione della domanda di lavoro e le caratteristiche del processo di digitalizzazione in funzione alla dimensione aziendale. Nella tabella 2 osserviamo che la quota media di assunzioni nel 2017 era piuttosto uniforme tra le varie classi dimensionali, mentre la sua crescita negli anni seguenti ha privilegiato relativamente le imprese medie (dal 18,2% al 25%) rispetto alle piccole (dal 17,6% al 22,3%) e a quelle con oltre 250 dipendenti (dal 18% al 22%). La dinamica positiva dei posti vacanti tra il 2018 e il 2022 sembra privilegiare, invece, le piccole (dal 2,1% al 4,4%) in comparazione alle aziende medio-grandi (dall'1,5% al 3,1%). Analoghe considerazioni valgono se esaminiamo il numero di posti vacanti calcolato rispetto al totale delle posizioni lavorative (occupate e vacanti).

Tabella 2. Domanda di lavoro per dimensione di impresa (valori medi)

	quota assunti		quota posti vacanti		quota posti vacanti*	
	T=2021	T-1=2017	T=2022	T-1=2018	T=2022	T-1=2018
[10-49]	22,3%	17,6%	4,4%	2,1%	3,5%	1,8%
[50-249]	24,9%	18,2%	3,1%	1,5%	2,6%	1,3%
[250-)	21,9%	17,7%	3,5%	2,0%	3,0%	1,7%
Totale	**22,7%**	**17,7%**	**4,2%**	**2,0%**	**3,4%**	**1,7%**

Nota: applicazione pesi longitudinali.

Fonte: elaborazioni degli Autori sul campione longitudinale RIL-2018 e RIL-2022

Tabella 3. Investimenti in nuove tecnologie per dimensione di impresa (valori medi)

	IA	Big Data	IoT e RA	Robotica
	T=2019-2021	T-1=2015-2017		
[10-49]	1,0%	5,9%	9,6%	4,7%
[50-249]	2,1%	13,7%	17,4%	12,6%
[250-)	10,6%	29,7%	39,9%	22,6%
Totale	**1,3%**	**7,4%**	**11,2%**	**6,1%**

Nota: applicazione pesi longitudinali.

Fonte: elaborazioni degli Autori sul campione longitudinale RIL-2018 e RIL-2022

Volgendo ora l'attenzione al processo di digitalizzazione, la tabella 3 rende evidente come tutte le tecnologie prese in esame nel periodo 2015-2017 erano allocate per una parte preponderante nelle grandi imprese mentre la loro incidenza era molto limitata nelle imprese con meno di 50 dipendenti. Questa sorta di dualismo tecnologico è chiaro soprattutto per Big Data e Cloud computing (5,9% vs 30%) e per l'Internet delle cose e Realtà aumentata (9,6% vs 40%) più che per Robotica e Smart manufacturing (4,7% vs 22,6%). Le tecnologie dell'informazione, e le connesse innovazioni organizzative, sono forti predittori dei futuri investimenti in tecnologie IA e, in questa prospettiva, la loro distribuzione tra le varie classi dimensionali aiuta a razionalizzare il fatto che la cosiddetta quarta rivoluzione industriale, per adesso, sembra riguardare quasi esclusivamente le grandi imprese.

3. Il modello econometrico

L'analisi econometrica arricchisce il quadro descrittivo nella misura in cui esamina la relazione tra processo di digitalizzazione, investimenti in tecnologie IA e la domanda di lavoro tenendo conto delle caratteristiche produttive, dell'assetto manageriale e della composizione della forza lavoro occupata. Viene quindi stimata la seguente equazione di regressione:

$$Y_{it} = \alpha + \beta_1 AI_{it=2019\text{-}21} + \beta_2 Dig\ Tech_{i,t\text{-}1=2015\text{-}17} + \beta_3 M_{it\text{-}1} + \beta_4 Y_{it\text{-}1} + \varepsilon_{it} \quad [1]$$

dove Y_{it} rappresenta alternativamente: i) la quota di lavoratori assunti con contratto di lavoro

dipendente nel corso del 2021 rispetto alla media degli occupati; ii) la quota di figure professionali che l'impresa dichiara di cercare nel 2022 rispetto alla media degli occupati; iii) la quota di posti vacanti nel 2022 rispetto al totale delle posizioni lavorative (occupate e vacanti) (Davis e Haltiwanger 1999 e 2014). La principale variabile esplicativa IA è una variabile dicotomica pari a 1 se l'impresa ha investito in tecnologie IA nel triennio 2019-2021 e pari a 0 altrimenti. Il processo di digitalizzazione è formalizzato dal vettore $Dig\ Tech_{it=2015\text{-}17}$ che include tre distinti indicatori per la pregressa adozione di tecnologie dell'informazione (Big Data e Internet delle Cose) e della produzione (Robotica) adottate nel corso del periodo 2015-2017. Il vettore M_{it-1}, rappresenta i controlli per l'assetto manageriale e la governance societaria, la composizione della forza lavoro, le caratteristiche produttive delle imprese, tra cui la specializzazione settoriale, la localizzazione geografica e altri indicatori di competitività (ricavi per dipendente, esposizione al commercio internazionale): tutte le variabili incluse in M_{it-1} sono calcolate sui dati dell'indagine RIL-2018. L'inclusione del valore ritardato della variabile dipendente – ossia la quota di assunzioni e di posti vacanti riferita alle annualità precedenti (2017 e 2018) – consente di controllare potenziali *bias* di stima derivanti da eterogeneità non osservata. Tale eterogeneità può essere associata a dinamiche preesistenti di capacità innovativa o di funzionamento dei mercati interni del lavoro, come ad esempio la diversa propensione, latente ma persistente, delle imprese a investire in innovazione o a ricercare personale qualificato rispetto ad altre con caratteristiche osservabili analoghe. Il parametro ε_{it} indica, infine, l'errore idiosincratico con media nulla e varianza finita.

L'equazione [1] viene stimata con semplici modelli di regressione lineare e attraverso l'applicazione di metodologie di *Propensity Score Matching* (PSM) (Rosenbaum e Rubin 1983) al fine di limitare ulteriormente il rischio di distorsione delle stime OLS. La nostra strategia empirica non risolve necessariamente tutti i potenziali problemi di autoselezione e di endogeneità connaturati agli investimenti in tecnologie IA e – conseguentemente – non permette un'interpretazione causale delle stime del coefficiente β_1. La formalizzazione dinamica dell'equazione [1] e l'applicazione dei modelli di regressione nel sottocampione di imprese con supporto comune nelle caratteristiche osservabili, tuttavia, consente un'interpretazione analitica dei risultati che va oltre la semplice correlazione multivariata (Dosi 2023).

4. I risultati

La tabella 4 riporta i risultati principali della stima dell'equazione [1] dove le misure di domanda di lavoro (quota di assunzioni e di posizioni vacanti) e l'investimento in IA sono riferiti ai dati RIL-2022, mentre l'insieme delle variabili di controllo sono calcolate sui dati dell'indagine RIL-2018. Le stime OLS riportate nelle colonne (1-3) suggeriscono che l'investimento IA non condiziona la futura evoluzione delle assunzioni (colonna 1) anche quando si tiene conto della pregressa adozione di tecnologie abilitanti (colonna 2) e dei potenziali nessi di causalità inversa tra domanda di lavoro e investimenti IA (colonna 3). I risultati riportati nelle rimanenti tre colonne mostrano, invece, che le tecnologie IA si accompagnano a un incremento seppur lieve della quota di posti vacanti, per un ordine di grandezza che varia tra +1% (colonna 4) e +0,7% (colonna 6), in base alla specificazione econometrica. Sebbene si tratti di variazioni contenute in valore assoluto, esse arricchiscono in modo sostanziale l'analisi discussa finora. In altre parole, la diffusione ancora marginale di ecosistemi IA nel sistema imprenditoriale italiano non produce una variazione effettiva della domanda di lavoro, ma sembra favorire comunque la ricerca di nuove figure professionali. La natura dei dati non permette di inferire la tipologia delle figure professionali e delle competenze che le imprese ricercano in esito degli investimenti in IA, sebbene alcune evidenze descrittive rivelino la prevalenza di profili altamente qualificati per l'espletamento di mansioni di tipo cognitivo e relazionale (Inapp 2024).

La tabella 5 riporta i risultati ottenuti dalla stima del modello *Propensity Score Matching* (PSM) che restringe l'analisi al sottocampione di imprese che appaiono del tutto analoghe dal punto di vista delle caratteristiche osservabili, ad eccezione della scelta di investire in IA riducendo così il *bias* di selezione. È possibile confermare i risultati precedenti: esiste una relazione positiva e statisticamente significativa tra l'investimento in tecnologie connesse all'IA e la domanda di nuovo personale dipendente (+1,2% e +1%), mentre non vi è correlazione con la quota di nuovi assunti.

Il fatto che i risultati ottenuti con il modello OLS siano coerenti con quelli ottenuti con il PSM supporta

Tabella 4. Stime OLS, variabile dipendente: quota nuovi assunti e quota posti vacanti

	Quota nuovi assunti			Quota posti vacanti		
	[1]	[2]	[3]	[4]	[5]	[6]
$IA_{2019-21}$	0,015	0,014	0,014	0,010**	0,009**	0,007*
	(0,019)	(0,019)	(0,016)	(0,004)	(0,004)	(0,004)
Big Data $_{2015-17}$		0,020	0,012		0,004*	0,003
		(0,013)	(0,011)		(0,002)	(0,002)
IoT e RA $_{2015-17}$		0,001	-0,000		-0,000	-0,001
		(0,011)	(0,009)		(0,002)	(0,002)
Robotica $_{2015-17}$		-0,017*	-0,007		0,001	0,002
		(0,009)	(0,008)		(0,002)	(0,002)
Quota assunti $_{(t-1)}$			0,504***			
			(0,035)			
Quota posti vacanti $_{(t-1)}$					0,259***	
						(0,042)
Altri controlli	Sì	Sì	Sì	Sì	Sì	Sì
Costante	0,235***	0,236***	0,140***	0,082***	0,083***	0,070***
	(0,045)	(0,045)	(0,041)	(0,011)	(0,011)	(0,011)
N. di osservazioni	7.343	7.343	7.343	7.343	7.343	7.343
R2	0,185	0,185	0,353	0,051	0,051	0,078

Nota: altri controlli includono – in valori ritardati – caratteristiche manageriali (istruzione, età e sesso di chi gestisce impresa, incidenza proprietà familiare), composizione della forza lavoro per livello professione, istruzione e genere, caratteristiche produttive delle imprese (export, incidenza riorganizzazione causa Covid, età in anni dalla fondazione, incidenza blocco licenziamenti causa Covid, (log del) fatturato per dipendente, (log del) numero dipendenti. Tutte le regressioni includono effetti fissi per settore di attività 2 digit, regione geografica. Errori Standard clusterizzati a livello di impresa sono riportati in parentesi. Significatività statistica: *** 1%, ** 5%, * 10%.

Fonte: elaborazioni degli Autori su dati RIL 2018-2022

Tabella 5. Stime OLS e Propensity Score Matching

	Quota nuovi assunti		Quota posti vacanti	
	[1]	**[2]**	**[3]**	**[4]**
IA $_{2019-21}$	0,006	0,016	0,012**	0,010*
	(0,023)	(0,020)	(0,005)	(0,006)
Big Data $_{205-17}$	-0,013	-0,002	0,015**	0,003
	(0,034)	(0,026)	(0,008)	(0,009)
IoT e RA $_{2015-17}$	-0,062	-0,047	-0,013	-0,016**
	(0,040)	(0,029)	(0,008)	(0,007)
Robotica $_{2015-17}$	0,017	0,018	0,006	0,005
	(0,027)	(0,023)	(0,007)	(0,007)
Quota assunti$_{(t-1)}$		0,480***		
		(0,141)		
Quota posti vacanti$_{(t-1)}$			0,195**	
				(0,080)
Altri controlli	Si	Si	Si	Si
Costante	0,181	-0,034	0,091*	0,001

segue

segue **Tabella 5**

	Quota nuovi assunti		Quota posti vacanti	
	[1]	[2]	[3]	[4]
	(0,211)	(0,192)	(0,046)	(0,040)
N. di osservazioni	468	468	464	464
R2	0,113	0,366	0,090	0,114

Nota: altri controlli includono – in valori ritardati – caratteristiche manageriali (istruzione, età e sesso di chi gestisce impresa, incidenza proprietà familiare), composizione della forza lavoro per livello professione, istruzione e genere, caratteristiche produttive delle imprese (export, incidenza riorganizzazione causa Covid, età in anni dalla fondazione, incidenza blocco licenziamenti causa Covid, (log) fatturato per dipendente, (log) numero dipendenti. Tutte le regressioni includono effetti fissi per settore di attività 2 digit, regione geografica. Errori Standard clusterizzati a livello di impresa sono riportati in parentesi. Significatività statistica: ***1%, **5%, *10%.

Fonte: elaborazione degli Autori su dati RIL 2018-2022

ancora una volta l'ipotesi che, almeno nel breve periodo, l'adozione di IA sembra favorire la ricerca di nuove figure professionali. In questa prospettiva – le nostre analisi sono coerenti con le evidenze secondo cui l'introduzione di tecnologie avanzate, come l'Intelligenza artificiale, possa generare nuove esigenze organizzative e operative, aumentando la domanda complessiva di lavoro (Acemoglu e Restrepo 2019).

Robustezze

Le evidenze discusse finora possono essere approfondite facendo riferimento a una diversa misura di domanda di lavoro e declinando le analisi in funzione della dimensione aziendale. In particolare, la tabella 6 illustra gli esiti delle regressioni quando la variabile dipendente dell'equazione [1] è definita dal numero di posti vacanti in rapporto al totale delle posizioni lavorative (occupati+posti vacanti). Si osserva ancora una volta come l'adozione di tecnologie IA si accompagna a una crescita, seppur lieve, del tasso di posti vacanti per un ammontare che può variare tra +0,8% (colonna 1) e +0,6% (colonna 3) in base alle diverse specificazioni. Il valore assoluto e la significatività di queste stime sono analoghi a quelle della tabella 4 e confermano come le argomentazioni precedenti siano robuste a diverse misure di domanda di lavoro potenziale.

La tabella 7 mostra gli esiti delle regressioni condotte separatamente per le piccole imprese

Tabella 6. Stime OLS, variabile dipendente: quota posti vacanti secondo definizione Eurostat

	[1]	[2]	[3]
IA $_{2019\text{-}21}$	0,008**	0,007**	0,006**
	(0,003)	(0,003)	(0,003)
Big Data $_{2015\text{-}17}$		0,004*	0,002
		(0,002)	(0,002)
IoT e RA $_{2015\text{-}17}$		0,000	-0,000
		(0,002)	(0,002)
Robotica $_{2015\text{-}17}$		0,001	0,001
		(0,002)	(0,002)
Quota posti vacanti$_{(t\text{-}1)}$			0,248***
			(0,031)
Altri controlli	Sì	Sì	Sì
Costante	0,067***	0,068***	0,059***
	(0,008)	(0,008)	(0,008)
N. di osservazioni	7.343	7.343	7.343
R2	0,055	0,055	0,082

Nota: Altri controlli includono – in valori ritardati – caratteristiche manageriali (istruzione, età e sesso di chi gestisce impresa, incidenza proprietà familiare), composizione della forza lavoro per livello professione, istruzione e genere, caratteristiche produttive delle imprese (export, incidenza riorganizzazione causa Covid, età in anni dalla fondazione, incidenza blocco licenziamenti causa Covid, (log) fatturato per dipendente, (log) numero dipendenti. Tutte le regressioni includono effetti fissi per settore di attività 2 digit, regione geografica. Errori Standard clusterizzati a livello di impresa sono riportati in parentesi. Significatività statistica: ***1%, **5%, *10%.

Fonte: elaborazione degli Autori su dati RIL 2018-2022

Tabella 7. Stime OLS per dimensione di impresa

	Quota assunti		Quota posti vacanti		Quota posti vacanti (Eurostat)	
	Piccole	Medie-grandi	Piccole	Medie-grandi	Piccole	Medie-grandi
	[1]	[2]	[3]	[4]	[5]	[6]
IA	0,066	-0,006	-0,007	0,011**	-0,006	0,009**
	(0,042)	(0,017)	(0,006)	(0,005)	(0,005)	(0,004)
Quota assunti $_{(t-1)}$	0,445***	0,569***				
	(0,051)	(0,045)				
Quota posti vacanti$_{(t-1)}$			0,252***	0,250***		
			(0,051)	(0,069)		
Quota posti vacanti (Eurostat) $_{(t-1)}$					0,237***	0,248***
					(0,038)	(0,050)
Altri controlli	Sì	Sì	Sì	Sì	Sì	Sì
Costante	0,084	0,202***	0,097***	0,047***	0,080***	0,039***
	(0,062)	(0,068)	(0,022)	(0,012)	(0,015)	(0,009)
N. di osservazioni	3.636	3.707	3.636	3.707	3.640	3.708
R2	0,317	0,399	0,070	0,090	0,074	0,095

Nota: altri controlli includono – in valori ritardati -caratteristiche manageriali (istruzione, età e sesso di chi gestisce impresa, incidenza proprietà familiare), composizione della forza lavoro per livello professione, istruzione e genere, caratteristiche produttive delle imprese (export, incidenza riorganizzazione causa Covid, età in anni dalla fondazione, incidenza blocco licenziamenti causa Covid, (log) fatturato per dipendente, (log) numero dipendenti. Tutte le regressioni includono effetti fissi per settore di attività 2 digit, regione geografica. Errori Standard clusterizzati a livello di impresa sono riportati in parentesi. Significatività statistica: ***1%, **5%, *10%.

Fonte: elaborazione degli autori su dati RIL 2018-2022

e per le realtà produttive con oltre 50 dipendenti. Si osserva che le tecnologie IA non influenzano la quota di assunzioni indipendentemente dalla dimensione aziendale, mentre quest'ultima appare fondamentale quando si considera l'intensità di ricerca di nuove figure professionali. Nella fattispecie, gli investimenti IA si associano a un incremento positivo del tasso di posti vacanti solo nelle imprese medio-grandi (+1,1% e + 0,9% nelle colonne 4 e 6, rispettivamente) a fronte di un effetto negativo e non significativo nelle piccole realtà produttive (colonne 3 e 5).

In sintesi, la relazione positiva tra ecosistemi IA e domanda potenziale di lavoro che abbiamo registrato per l'intera economia sembra trainata esclusivamente da ciò che avviene nelle aziende con oltre 50 dipendenti, ovvero non riguarda la parte maggioritaria delle imprese del nostro Paese. La tabella 6 produce dunque evidenze coerenti con quanto delineato nelle statistiche descrittive e conferma l'ipotesi che la dinamica delle nuove tecnologie è fortemente condizionata dalla forte eterogeneità del tessuto produttivo che si manifesta in modo sostanziale, sebbene non esclusivo, nel dualismo dimensionale e competitivo delle imprese.

Conclusioni

La ricerca sviluppata nelle pagine precedenti ha analizzato la relazione tra l'adozione di tecnologie digitali, ed in particolare gli investimenti in sistemi di Intelligenza artificiale (IA) di prima generazione, e la domanda di lavoro (effettiva e potenziale) espressa dalle imprese italiane.

Le elaborazioni hanno permesso di illustrare alcuni risultati relativamente inediti per il caso italiano. Innanzitutto, l'adozione di sistemi IA nel corso del periodo 2019-2021 coinvolge in media solo l'1% delle imprese, mentre l'incidenza delle tecnologie digitali ad essi complementari appare relativamente più elevata: Big Data 7%, Internet delle cose 11%, Robotica 6%. Secondo, l'investimento in IA in sé non esercita una influenza significativa sulla quota di nuove assunzioni (domanda effettiva) mentre, l'investimento in IA considerato in aggiunta all'adozione di tecnologie digitali avanzate, si associa

ad un incremento positivo della quota di posti vacanti nelle imprese (+1%). Nonostante le tecniche di analisi e la natura dei dati non consentano di stabilire una relazione causale, si possono avanzare alcune argomentazioni a fini di policy.

Tali evidenze suggeriscono, che l'efficacia dell'IA nel generare domanda di lavoro dipende dalla complementarietà con le traiettorie tecnologiche pregresse che, a sua volta, interagisce con una molteplicità di altri fattori microeconomici. L'Intelligenza artificiale, infatti, tipicamente, non è implementata come tecnologia autonoma, ma come parte di un ecosistema digitale più ampio che include, ad esempio, cloud computing, Internet of Things (IoT), Big Data analytics o sistemi ERP intelligenti. Queste tecnologie, se integrate in modo funzionale nella vita aziendale, possono riorganizzare i processi produttivi e gestionali in modo radicale, ampliando le strategie competitive e rendendo necessaria l'acquisizione di nuove figure professionali e competenze, soprattutto per quei lavoratori occupati in professioni complementari al nuovo paradigma. La diffusione dell'IA porta con sé anche la questione fondamentale dei flussi in uscita dal mercato (imprese e lavoratori), ovvero dell'effetto netto che un'applicazione pervasiva delle nuove tecnologie potrà avere sull'occupazione, sui salari e su vari aspetti dello sviluppo economico nel medio periodo. La nostra analisi non ha trattato questo aspetto del fenomeno che, peraltro, avrebbe chiamato in causa una riflessione più generale di politica pubblica che andava oltre le finalità del presente articolo. Future linee di ricerca avranno l'obiettivo di approfondire proprio tali questioni.

Bibliografia

Acemoglu D., Anderson G., Beede D., Buffington C., Childress E., Dinlersoz E., Foster L., Goldschlag N., Haltiwanger J., Kroff Z., Restrepo P., Zolas N. (2023a), Advanced technology adoption: Selection or causal effects?, *AEA Papers and Proceedings*, 113, pp.210-214

Acemoglu D., Koster H.R.A., Ozgen C. (2023b), *Robots and workers: Evidence from the Netherlands*, Working Paper n.31009, Cambridge, National Bureau of Economic Research

Acemoglu D., Restrepo P. (2020), Robots and jobs: Evidence from US labor markets, *Journal of Political Economy*, 128, n.6, pp.2188-2244

Acemoglu D., Restrepo P. (2019), Artificial intelligence, automation, and work, in Agrawal A., Gans J., Goldfarb A. (eds.), *The Economics of Artificial Intelligence: An Agenda*, Chicago, The University of Chicago Press, pp.197-236

Acemoglu D., Restrepo P. (2018), The race between man and machine: Implications of technology for growth, factor shares, and employment, *American Economic Review*, 108, n.6, pp.1488-1542

Acemoglu D., Autor D., Hazell J., Restrepo P. (2022), Artificial intelligence and jobs: evidence from online vacancies, *Journal of Labor Economics*, 40, n.S1, pp.S293-S340

Agrawal A.K., Gans J.S., Goldfarb A. (2023), Artificial intelligence adoption and system-wide change, *Journal of Economics and Management Strategy*, 33, n.2, pp.327-337

Agrawal A.K., Gans J.S., Goldfarb A. (2022), *Power and Prediction: The Disruptive Economics of Artificial Intelligence*, Boston, Harvard Business Press

Agrawal A.K., Gans J.S., Goldfarb A. (2018), *Prediction Machines: The Simple Economics of Artificial Intelligence*, Boston, Harvard Business Review Press

Agrawal A.K., Gans J.S., Goldfarb A. (eds.) (2019), *The Economics of Artificial Intelligence: An Agenda*, Cambridge, National Bureau of Economic Research

Aghion P., Antonin C., Bunel S., Jaravel X. (2020), *What are the labor and product market effects of automation? New evidence from France*, CEPR Discussion Paper n.14443, London, Centre for Economic Policy Research

Albanesi S., da Silva A.D., Jimeno J.F., Lamo A., Wabitsch A. (2025), New technologies and jobs in Europe, *Economic Policy*, 40, n.121, pp.71-139

Alekseeva L., Azar J., Giné M., Samila S., Taska B. (2021), The demand for AI skills in the labor market, *Labour Economics*, 71, art.102002

Autor D., Levy F., Murnane R.J. (2003), The skill content of recent technological change: An empirical exploration, *The Quarterly Journal of Economics*, 118, n.4, pp.1279-1333

Autor D., Salomons A. (2018), Is automation labor share–displacing? Productivity growth, employment, and the labor share, *Brookings Papers on Economic Activity*, 1, pp.1-87

Babina T., Fedyk A., He A., Hodson J. (2024), Artificial intelligence, firm growth, and product innovation, *Journal of Financial Economics*, 151, art.103745

Battisti M., Brunetti I., Gravina A.F., Li Donni P. (2024), *Automation and young workers'job trajectories: The Italian case*, Inapp Working Paper n.124, Roma, Inapp

Bessen J. (2017), *Automation and jobs: When technology boosts employment*, Law and Economics Research Paper n.17-09, Boston, Boston University School of Law

Bessen J., Goos M., Salomons A., van den Berge W. (2023), What happens to workers at firms that automate?, *The Review of Economics and Statistics*, 107, n.1, pp.125-141

Bonfiglioli A., Crinò R., Fadinger H., Gancia G. (2023), Robot imports and firm-level outcomes, *Economic Journal*, 134, n.664, pp.3428-3444

Bonfiglioli A., Gancia G., Papadakis I., Crinò R. (2025), Artificial intelligence and jobs: Evidence from US commuting zones, *Economic Policy*, 40, n.121, pp.145-194

Bostrom N., Yudkowsky E. (2014), The ethics of artificial intelligence, in Frankish K., Ramsey W.M. (eds.), *The Cambridge Handbook of Artificial Intelligence*, Cambridge, Cambridge University Press, pp.316-334

Bratta B., Romano L., Acciari P., Mazzolari F. (2023), Assessing the impact of digital technology diffusion policies: Evidence from Italy, *Economics of Innovation and New Technology*, 32, n.8, pp.1114-1137

Brunetti I., Dughera S., Ricci A. (2025), Technological path-dependency in AI adoption: Evidence from Italian firms, *Economics of Innovation and New Technology*, in corso di pubblicazione

Brynjolfsson E., Chandar B, Chen R. (2025), *Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence*, Stanfors University <https://digitaleconomy.stanford.edu/wp-content/uploads/2025/08/Canaries_BrynjolfssonChandarChen.pdf>

Brynjolfsson E., McAfee A. (2014), *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, New York, W.W. Norton & Co

Brynjolfsson E., Mitchell T., Rock D. (2018), What can machines learn, and what does it mean for occupations and the economy? *AEA Papers and Proceedings*, 108, pp.43-47

Brynjolfsson E., Rock D., Syverson C. (2021), The Productivity J-Curve: How Intangibles Complement General Purpose Technologies, *American Economic Journal: Macroeconomics*, 13, n.1, pp.333-372

Calvino F., Fontanelli L. (2024), *AI users are not all alike: The characteristics of French firms buying and developing AI*, CESifo Working Paper Series n.11466, München, CESifo

Calvino F., Fontanelli L. (2023), *A portrait of AI adopters across countries: Firm characteristics, assets'complementarities and productivity*, OECD Science, Technology and Industry Working Papers n.2023/02, Paris, OECD Publishing

Caselli M., Fourrier-Nicolai E., Fracasso A., Scicchitano S. (2024), *Digital technologies and firms'employment and training*, CESifo Working Paper n.11056, München, CESifo

Cho J., DeStefano T., Kim H., Kim I., Paik J.H. (2022), What's driving the diffusion of next-generation digital technologies?, *Technovation*, 119, 102477 <https://doi.org/10.1016/j.technovation.2022.102477>

Ciani E., Lattanzio S., Graziella Mendicino G., Viviano E. (2025), *L'Occupazione in Italia dopo la pandemia*, Questioni di Economia e Finanza n.962, Roma, Banca d'Italia Cockburn I., Henderson R., Stern S. (2019), The impact of artificial intelligence on innovation, in Agrawal A., Gans J., Goldfarb A. (a cura di), *The Economics of Artificial Intelligence: An Agenda*, Chicago, University of Chicago Press

Czarnitzki D., Fernández G.P., Rammer C. (2023), Artificial intelligence and firm-level productivity, *Journal of Economic Behavior & Organization*, 211, pp.188-205

Dalla Zuanna A., Dottori D., Gentili E., Lattanzio S. (2024), *An assessment of occupational exposure to artificial intelligence in Italy*, Questioni di Economia e Finanza n.878, Roma, Banca d'Italia

Davis S., Haltiwanger J. (2014), Labor market fluidity and economic performance, *NBER Working Paper* n.20479, Cambridge, National Bureau of Economic Research

Davis S., Haltiwanger J. (1999), Gross job flows, in Ashenfelter O., Card D. (a cura di), *Handbook of Labor Economics*, 3B, Amsterdam, North-Holland

Dosi G. (2023), *The Foundations of Complex Evolving Economies*, Oxford, Oxford University Press

Draca M., Nathan M., Nguyen-Tien V., Oliveira-Cunha J., Rosso A., Valero A. (2024), *The new wave? The role of human capital and STEM skills in technology adoption in the UK*, IZA Discussion Paper n.17329, Bonn, IZA

Drydakis N. (2025), *Artificial intelligence and labour market outcomes*, IZA World of Labor n.514, Bonn, IZA

Eloundou T., Manning S., Mishkin P., Rock D. (2023), Gpts are gpts: An early look at the labour market impact potential of large language models, *arXiv preprint* <https://doi.org/10.48550/arXiv.2303.10130>

Felten E., Raj M., Seamans R. (2023), *Occupational heterogeneity in exposure to generative AI*, SSRN Working Paper n.4414065 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4414065>

Felten E., Raj M., Seamans R. (2021), Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses, *Strategic Management Journal*, 42, n.12, pp.2195-2217

Graetz G., Michaels G. (2018), Robots at Work, *The Review of Economics and Statistics,* 100, n.5, pp.753-768

Guarascio G., Reljic J. (2025), AI and employment in Europe, *Economics Letters*, 247, art.111531

Igna I., Venturini F. (2023), The determinants of AI innovation across European firms, *Research Policy*, 52, n.2, art.104661

Inapp (2024), *Rapporto Inapp 2024. Lavoro e formazione: necessario un cambio di paradigma*, Roma, Inapp

Juhász R., Squicciarini M., Voigtländer N. (2024), Technology adoption and productivity growth: Evidence from industrialization in France, *Journal of Political Economy*, 132, n.10, pp.3215-3259

Kogan L., Papanikolaou D., Seru A., Stoffman N. (2017), Technological innovation, resource allocation, and growth, *The Quarterly Journal of Economics*, 132, n.2, pp.665-712

Noy S., Zhang W. (2023), Experimental evidence on the productivity effects of generative artificial intelligence, *Science*, 381, n.6654, pp.187-192

OECD (2019), *Artificial Intelligence in Society*, Paris, OECD Publishing

OECD (2023), *OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market*, Paris, OECD Publishing

Pizzinelli C., Panton A., Mendes Tavares M., Cazzaniga M., Li L. (2023), *Labour market exposure to AI: Cross-country differences and distributional implications*, IMF Working Paper n.2023/216, Washington, International Monetary Fund

Schwab K. (2025), *The Fourth Industrial Revolution*, London, Portfolio Penguin

Zolas N., Kroff Z., Brynjolfsson E., McElheran K., Beede D. N., Buffington C., Goldschlag N., Foster L., Dinlersoz E. (2020), *Advanced Technologies Adoption and Use by U.S. Firms: Evidence from the Annual Business Survey, National Bureau of Economic Research Working Papers* n.28290, Cambridge (MA), NBER

Webb M. (2019), *The impact of artificial intelligence on the labour market*, SSRN Working Paper n.3482150 <http://dx.doi.org/10.2139/ssrn.3482150>

Irene Brunetti
i.brunetti@inapp.gov.it
Ricercatrice in economia applicata presso l'Inapp e responsabile scientifica di un accordo di collaborazione tra Inapp e il Dipartimento di Economia "M. Biagi" dell'Università di Modena e Reggio-Emilia. I suoi interessi di ricerca riguardano le dinamiche del mercato del lavoro, la valutazione delle politiche attive e la mobilità socioeconomica. È autrice di ricerche pubblicate su riviste nazionali e internazionali tra cui: Even more discouraged? The NEET generation at the age of Covid-19, *Applied Economics*, 2024; Occupational mobility: Theory and estimation for Italy, *The Journal of Economic Inequality*, 2023

Andrea Ricci
an.ricci@inapp.gov.it
Dirigente di ricerca in economia ed economia applicata presso l'Inapp, dove coordina la Struttura di ricerca Imprese, transizioni e sviluppo del Mezzogiorno. I suoi principali interessi di ricerca hanno per oggetto la valutazione delle politiche per il lavoro, l'economia del mercato del lavoro, l'organizzazione industriale e il comportamento delle imprese. È autore di ricerche pubblicate su riviste nazionali e internazionali.

Determinare il mismatch delle competenze usando la GenAI

Costruzione di un modello su istruzione, formazione e mercato del lavoro in Italia

Alessia Romito
INAPP

Boris Sofronic
INAPP

Lo studio analizza l'utilizzo dei *Large Language Models* (LLM) per l'individuazione del mismatch delle competenze in Italia. Confrontando due sperimentazioni condotte nel 2023 e nel 2025, evidenzia l'evoluzione dei modelli di Intelligenza artificiale generativa, dai limiti iniziali (errori fattuali, obsolescenza della knowledge-base, allucinazioni) ai recenti progressi (Chain-of-Thought, RLHF, RAG). Il lavoro identifica metodologie efficaci per l'impiego rigoroso dei LLM nella ricerca: selezione accurata delle fonti, personalizzazione mirata, prompt-engineering strategico e approccio ibrido con supervisione umana.

This study analyses the use of Large Language Models (LLMs) to identify skills mismatch in Italy. By comparing two experiments conducted in 2023 and 2025, it traces the evolution of generative AI models, from early limitations (factual inaccuracies, obsolescence of knowledge bases, hallucinations) to recent advancements such as Chain-of-Thought prompting, Reinforcement Learning from Human Feedback (RLHF), and Retrieval-Augmented Generation (RAG). The study identifies effective methodologies for rigorous LLM application in research: accurate source selection, targeted customisation, strategic prompt engineering, and a hybrid approach involving human oversight.

DOI: 10.53223/Sinappsi_2025-02-14

Citazione
Romito A., Sofronic B. (2025), Determinare il mismatch delle competenze usando la GenAI. Costruzione di un modello su istruzione, formazione e mercato del lavoro in Italia, *Sinappsi*, XV, n.2, pp.208-227

Parole chiave
Intelligenza artificiale
Mercato del lavoro
Skill mismatch

Keywords
Artificial intelligence,
Labour market
Skills mismatch

Introduzione[1]

I modelli linguistici di grandi dimensioni (*large language models*, LLM) sono modelli di Intelligenza artificiale (IA) generativa (GenAI) che utilizzano l'apprendimento automatico (*machine learning*, ML) per comprendere e generare linguaggio umano. Sono basati su reti neurali artificiali (*artificial neural networks*, ANN) e tecniche di elaborazione del linguaggio naturale (*natural language processing*, NLP). A marzo 2025 i LLM più conosciuti e più usati sono

1 Alla fine del documento è presente il glossario dei termini più tecnici.

i *Generative Pre-trained Transformer* (GPT) come: ChatGPT della OpenAI, Microsoft Copilot, Claude.ai della Antrophic, DeepSeek, Google Gemini, Llama della Meta, Grok della xAI, Mistral AI e molti altri. Questi modelli si basano su un'architettura chiamata *Transformer*, un'innovazione fondamentale nel campo delle ANN, introdotta da ricercatori di Google Brain e Google Research nel paper *Attention Is All You Need* (Vaswani *et al.* 2017). I Transformer sono progettati per analizzare le relazioni e le dipendenze tra gli elementi di una sequenza, come le parole in una frase, utilizzando meccanismi di *Self-attention* per determinare l'importanza di ogni elemento e generare risposte sempre più accurate (*ibidem*, 3).

Le sperimentazioni sui LLM da parte della comunità scientifica internazionale hanno prodotto risultati misti. Nonostante gli entusiasmanti progressi offerti dai LLM, la letteratura scientifica documenta anche limiti significativi; ad esempio, sono stati osservati errori fattuali (Augenstein *et al*. 2023), allucinazioni artificiali (Alkaissi e McFarlane 2023; Wang *et al.* 2023), errori causati dai dati reperiti da fonti esterne o errori di inferenza (Wang *et al.* 2023, 5)[2]. Questo testo descrive alcune esperienze di utilizzo sperimentale dei LLM nell'individuazione del mismatch delle competenze e, più in generale, l'uso dei LLM nei domini dell'istruzione, della formazione, delle professioni e dell'occupazione condotte in Inapp e ha come obiettivo l'identificazione delle metodologie per rilevare errori fattuali e allucinazioni, analizzarne le cause e sviluppare strategie per un uso consapevole dell'IA nella ricerca scientifica. Il metodo è stato già sperimentato da Chen *et al.* (2024) seppur con finalità differenti.

Come vedremo di seguito, l'IA consiste di un ampio panorama di tecnologie progettate per consentire alle macchine di percepire, interpretare, apprendere e agire emulando le capacità cognitive umane. All'interno di questa gamma delle soluzioni IA, i LLM e la GenAI occupano un posto centrale, essendo in grado di creare contenuti originali, dai testi alle immagini, fino ad audio e video. Altri modelli IA, invece, si concentrano su compiti specifici, come il riconoscimento di pattern, ottimizzando processi ripetitivi per aumentare la produttività piuttosto che creare nuovi contenuti (Cazzaniga *et al*. 2024). In attesa della cosiddetta Intelligenza artificiale generale (AGI), tecnologie più tradizionali e mature, come i sistemi di ML basati sulle librerie open source come PyTorch (Ansel *et al*. 2024) e TensorFlow[3] o piattaforme commerciali come IBM Watson Studio, Google Cloud AutoML e Microsoft Azure Machine Learning Studio, continuano a dimostrarsi più utili per assistere la ricerca scientifica (Singh 2023). Questo è particolarmente vero nei settori dell'istruzione e dell'occupazione, dove la qualità dei dati e l'accuratezza dell'analisi sono fondamentali.

1. I confini regolamentari e istituzionali della sperimentazione

Le tecnologie basate sull'Intelligenza artificiale stanno diventando sempre più pervasive, mostrando il loro potenziale trasformativo sulle dinamiche sociali, quali ad esempio la formazione, l'occupazione, la ricerca, nonché su quelle produttive. A fronte di una indispensabile necessità di accogliere l'innovazione, è emerso un altrettanto imprescindibile bisogno di regolamentarne l'uso, allo scopo di garantire che i sistemi di IA siano sicuri, etici e affidabili e che siano sviluppati e utilizzati in modo responsabile.

Il Consiglio europeo ha fatto un passo decisivo in questa direzione adottando il *Regolamento europeo sull'Intelligenza Artificiale*[4] (AI Act), entrato in vigore il 1° agosto 2024. Il documento affronta i potenziali rischi per la salute, la sicurezza e i diritti fondamentali dei cittadini, fornisce requisiti e obblighi a sviluppatori e operatori per quanto riguarda gli usi

2 Si veda la figura 1 del presente articolo.

3 TensorFlow Developers, *TensorFlow: Large-scale machine learning on heterogeneous systems* (Computer software), https://doi.org/10.5281/zenodo.4724125.

4 Parlamento europeo, Consiglio dell'Unione europea (2024), *Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (Regolamento sull'Intelligenza artificiale)*, https://eur-lex.europa.eu/eli/reg/2024/1689/oj.

specifici dell'IA e introduce un quadro uniforme in tutti i Paesi dell'UE basato sul rischio[5].

A livello nazionale, un primo modello di governance dell'IA in Italia viene elaborato a maggio del 2024, quando il Dipartimento per il Programma del Governo della Presidenza del Consiglio dei ministri pubblica un focus in materia di IA che analizza l'AI Act e presenta un Disegno di Legge (DDL) per l'approvazione parlamentare[6]; nel medesimo periodo il Dipartimento per la Trasformazione digitale e l'Agenzia per l'Italia digitale (Agid) pubblicano il documento integrale della *Strategia Italiana per l'Intelligenza artificiale 2024-2026* (Agid 2024). Il documento pone particolare attenzione all'implementazione dell'IA nella ricerca, nella pubblica amministrazione (PA), nelle imprese e nella formazione, definendo obiettivi e azioni strategiche per ciascun ambito. Per quanto riguarda la ricerca scientifica, l'Agid suggerisce che l'Italia dovrà 'rafforzare gli investimenti nell'IA, promuovere la creazione di competenze di ricerca e tecnologie calate nel contesto del Sistema-Paese in linea con principi di affidabilità e responsabilità, propri ai valori antropocentrici dell'UÈ, rimarcando la 'libertà di sperimentazione, utilizzando contenuti e dati per la creazione di dataset e l'addestramento di modelli resi disponibili in open source' (*ibidem*, 9).

Tra le strategie contemplate nella macroarea ricerca, coerenti con le finalità dello studio presentato, si segnalano: "il consolidamento dell'ecosistema italiano della ricerca, la progettazione di LLM italiani, progetti interdisciplinari per il benessere sociale, la ricerca fondazionale e blu-sky per l'IA di prossima generazione" (*ibidem*, 14).

È all'interno di questo perimetro, delineato da un quadro giuridico, europeo e nazionale, e di indirizzo, che l'Inapp è stato chiamato ad agire da parte del Ministero del Lavoro[7]. Nel documento programmatico *Indirizzi strategici triennio 2025-2027*[8], tra le tematiche di rilevanza strategica, viene promossa "l'analisi e la valutazione degli impatti delle tecnologie digitali in generale (e più nello specifico dell'Intelligenza artificiale) sull'occupazione, sui fabbisogni di aggiornamento e di riconversione delle competenze […]". In questo contesto si inseriscono le sperimentazioni sull'IA già svolte nell'ambito delle attività dell'Istituto con lo scopo di fornire spunti utili a creare una base di partenza per le attività future che l'ente intenderà avviare in questa direzione e condividere con la comunità di ricerca informazioni sui risultati raggiunti e criticità riscontrate. Un'esperienza senz'altro interessante è stato il progetto Sophia in cui l'Inapp ha sperimentato l'implementazione di diverse soluzioni IA in una piattaforma di Labour Market Intelligence (LMI), con l'intento di misurare e segnalare il mismatch delle competenze analizzando le *web job vacancies*, risultati di apprendimento delle opportunità formative e CV degli utenti (Sofronic 2022). Il progetto mirava a interconnettere le fonti dati sulle opportunità di apprendimento e quelle sulla domanda del mercato del lavoro usando le tecniche di ML, NLP e gli algoritmi di IA che riescono a collegare direttamente i concetti estratti dai testi destrutturati, ad es. Named Entity Recogniton (NER) o l'IA cognitiva. L'output desiderato era un prototipo funzionante di un LMI in grado di rispondere alle domande sulle competenze, professioni e *soft skills* più richieste, i settori economici più dinamici, le qualificazioni più spendibili per la crescita professionale, le tendenze a medio termine e molto altro. Prima di concentrarsi sulle successive fasi di sviluppo del progetto, è stato necessario e doveroso effettuare un'analisi preliminare accurata, mirata a verificare empiricamente l'affidabilità dei responsi generati dai modelli IA allora disponibili, rispettando i criteri del rigore scientifico.

5 Il Regolamento stabilisce quattro livelli di rischio e ne definisce specifiche regole (https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai):
- rischio minimo: i filtri spam e i videogiochi che sfruttano l'AI non sono soggetti ad alcun obbligo ai sensi del Regolamento;
- rischio specifico per la trasparenza: i sistemi come le chatbot devono informare chiaramente gli utenti che stanno interagendo con una macchina, mentre alcuni contenuti generati dall'IA devono essere etichettati come tali;
- rischio alto: i software medici basati sull'AI o i sistemi di AI utilizzati per il reclutamento, devono rispettare requisiti rigorosi, tra cui sistemi di mitigazione del rischio, set di dati di alta qualità, informazioni chiare sugli utenti, supervisione umana ecc;
- rischio inaccettabile: i sistemi di AI che permettono l'attribuzione di un 'punteggio sociale' da parte di governi o imprese sono considerati una chiara minaccia per i diritti fondamentali delle persone e sono pertanto vietati.

6 Atto Senato n. 1146 si veda https://www.senato.it/leggi-e-documenti/disegni-di-legge/scheda-ddl?did=58262.

7 Cfr. *Indirizzi strategici triennio 2025-2027*, Inapp (Delibera CdA 25 novembre 2024, n. 18).

8 Inapp, Delibera CdA 25 novembre 2024, n. 18, cit.

2. I risultati della prima sperimentazione

Il 30 novembre del 2022 entra in produzione la prima versione pubblica di ChatGPT che, con cento milioni di utenti attivi al mese, dopo un solo bimestre dal lancio, si impone come l'applicazione in più rapida crescita della storia (Augenstein *et al*. 2023). Grazie alle recensioni all'epoca molto positive da parte degli esperti del settore, alla facilità di accesso e alla possibilità di utilizzo limitato gratuito, da gennaio del 2023 il ChatBot di OpenAI si afferma come principale piattaforma di verifica empirica della qualità delle risposte generate da un LLM. Tali considerazioni hanno incoraggiato la sperimentazione degli strumenti di GenAI anche a supporto del modello di Sophia, già basato sulle tecnologie di IA mature e affermate, come ML e NLP.

Nel box 1[9] si riporta una prima interazione con il ChatGPT mirata a verificare in che misura la sua epistemologia fosse aggiornata alle nuove competenze richieste, alle professioni emergenti e alla disponibilità delle rispettive opportunità di apprendimento utili al reskilling e upskilling.

I risultati di questa prima conversazione sembravano incoraggianti e promettenti. Le risposte generate da ChatGPT risultavano logiche, coerenti e ben elaborate. Una breve verifica a campione ha confermato l'esistenza all'epoca di alcuni corsi di laurea, master e dottorato menzionati dal modello.

Le successive interazioni con il modello hanno prodotto risultati misti. Nelle risposte fornite sono stati riscontrati alcuni errori, talvolta critici. Alcuni autori li definiscono come *errori fattuali* (Augenstein *et al.* 2023), alcuni come *allucinazioni artificiali* (Alkaissi e McFarlane 2023), altri invece hanno preferito approfondire l'argomento e distinguere gli errori di risposta in diverse categorie con lo scopo di individuare meglio le cause e proporre eventuali rimedi (Wang *et al*. 2023). Nel paper *Survey on Factuality in Large language models: knowledge, retrieval and domain-specificity* (*ibidem*, 5), gli autori definiscono tutti gli errori riscontrati come fattuali, ma propongono una distinzione tra errori causati al livello del LLM, errori generati dai dati reperiti da fonti esterne ed errori di inferenza (figura 1).

Risulta particolarmente utile concentrare l'attenzione sulla prima e sulla terza categoria (gli errori causati al livello del modello e gli errori causati dall'inferenza), in quanto in buona parte corrispondono all'esperienza riscontrata durante le sperimentazioni all'epoca condotte in Inapp. La ricerca ha, infatti, rilevato tre tipologie principali di errori nelle risposte del

Figura 1. Esempi di diversi tipi di errori fattuali generati dai LLM

Category	Cause	Example Dialog	Notes and references
Model-level causes	Domain knowledge deficit	Q: CEO of Assicurazioni Generali? **BloombergGPT:** Philippe Donnet **GPT-NeoX:** Antonio De Lorenzo, Simone Gambarini, Enrico Zanetti **FLAN-T5-XXL:** John M Forsyth, Christopher K Peters, [empty string]	BloombergGPT is a finance domain-specific language model. [278]
	Outdated information	Q: When was Kyiv attacked by Russia? **ChatGPT:**As of my last knowledge update in September 2021, Russia had not launched an attack on Kyiv.	Kyiv was attacked by Russia on 25 February 2022.
	Reasoning error	Q: Who is Tom Cruise's mother? A: Mary Lee Pfeiffer Q: Who is Mary Lee Pfeiffer's son? A: There is no widely known information about...	From Berglund et al. [13]. It is clear that the model knows Tom Cruise's mother is Lee Pfeiffer, but it fails to reason that Lee Pfeiffer has a son named Tom Cruise.
Inference-level causes	Snowballing	Q: Was there ever a US senator who represented the state of New Hampshire and whose alma mater was the University of Pennsylvania? A: Yes... His name was John P. Hale	John P. Hale was graduated from Bowdoin College. [301]
	Exposure bias	Q: Aamari was very stoic. [PRONOUN] rarely showed any emotion. A: He.	The correct answer was Xe according to [96].

Fonte: Wang *et al*. (2023)

9 I box (da 1 a 9) sono consultabili a fine articolo.

modello: 1) scarsa conoscenza del dominio, 2) obsolescenza della knowledge-base, 3) allucinazioni parziali costruite usando elementi veri (errori di inferenza), di seguito esaminati.

La scarsa conoscenza del dominio è stata riscontrata in una delle interazioni con il ChatGPT – di cui un estratto è riportato nel box 2 – che aveva come obiettivo la verifica della conoscenza dei principali sistemi classificatori necessari per individuare meglio il mismatch delle competenze nei diversi settori economici e le professioni.

Il ChatBot ha generato una risposta errata rispetto alla conoscenza della classificazione Istat CP2011, fondamentale per le finalità della sperimentazione. Pur ammettendo di aver fornito un'informazione distorta e accettando la 'correzione', il modello risultava quindi essere poco accurato e, di conseguenza, poco affidabile ai fini della ricerca.

Considerando tuttavia l'evoluzione tecnologica, il continuo apprendimento del ChatBot e in particolar modo facendo affidamento al prompt-engineering[10], a distanza di circa un anno e mezzo, si è ritenuto doveroso effettuare la medesima interrogazione, allo scopo di verificare se il modello fosse in grado di generare un responso corretto alla stessa domanda. L'interrogazione ha prodotto un esito alquanto positivo, la risposta generata infatti mostra elementi di maggior accuratezza (box 3) e conduce a ritenere plausibile che l'interattività tra il modello e l'essere umano può contribuire all'evoluzione del LLM.

Rispetto alla obsolescenza della knowledge-base, si riporta una conversazione con ChatGPT relativa alle eventuali opportunità formative, presenti nel territorio della Provincia di Roma, per acquisire la qualificazione di estetista (box 4), il cui profilo è regolamentato nel Repertorio nazionale dell'istruzione e della formazione professionale (Operatore del benessere - Erogazione dei servizi di trattamento estetico).

Lo scopo della conversazione era quello di verificare se l'epistemologia del modello interrogato permetteva l'interconnessione tra i dati sui centri di formazione accreditati (CFP) censiti da Inapp nel Portale Accreditamento, le qualificazioni contenute nell'Atlante del Lavoro e delle qualificazioni e le rispettive opportunità formative disponibili per le iscrizioni ai corsi da erogare nel prossimo futuro utili a conseguire tali qualificazioni se presenti nei siti web dei singoli enti. La risposta generata dal LLM presentava informazioni errate. Rispetto alle allucinazioni artificiali, occorre fare riferimento alla fase di realizzazione di una delle azioni previste nel progetto Sophia, nello specifico l'attività di analisi delle piattaforme informatiche per l'istruzione e la formazione professionale in Europa e lo studio per una loro interoperabilità.

Si è ritenuto interessante 'mettere alla prova' il LLM, la cui sperimentazione sull'accuratezza proseguiva in parallelo, per individuare le fonti scientifiche e accademiche sui temi d'interesse e in linea con i contenuti trattati, interrogando il modello sulla richiesta di riferimenti bibliografici relativi all'argomento ICT tools on E&T in Europe. Gli esiti dell'interrogazione, riportati nel box 5, sono stati verificati attraverso il motore di ricerca dedicato (Google Scholar) e attraverso specifici strumenti di ricerca di documenti di natura accademica e scientifica; il risultato ottenuto dal confronto ha evidenziato il fenomeno delle allucinazioni artificiali poiché nessun riferimento bibliografico ha trovato corrispondenza in una fonte reale.

La prima sperimentazione sull'affidabilità e l'accuratezza dei LLM, condotta prevalentemente nella prima parte del 2023 usando ChatGPT – unico strumento di GenAI disponibile all'epoca – ha evidenziato diverse criticità (scarsa conoscenza del dominio, obsolescenza della *knowledge-base* e allucinazioni parziali) che meritano un'attenta riflessione. Le risposte generate da ChatGPT riguardo a sistemi classificatori come la CP2011 hanno mostrato una comprensione inizialmente errata del dominio, sebbene successivamente corretta. Questo miglioramento tra interrogazioni distanziate nel tempo suggerisce un possibile aggiornamento della knowledge-base, ma non garantisce l'assenza di errori in altri ambiti specialistici.

La problematica dell'obsolescenza delle informazioni, evidenziata nel caso dell'ente di formazione non più attivo, rappresenta un limite significativo per l'utilizzo di questi strumenti in contesti dinamici come quello delle opportunità formative e del mercato del lavoro. Il fenomeno delle 'allucinazioni artificiali' si è manifestato in modo particolarmente evidente nel caso dei riferimenti bibliografici errati,

10 L'arte di formulare domande a un LLM in modo efficiente per ottimizzare le risposte in funzione delle nostre esigenze (Bolognesi 2024).

che non sono risultati rintracciabili neppure attraverso strumenti di ricerca specialistici. Questo aspetto solleva interrogativi fondamentali sull'affidabilità dei LLM in contesti che richiedono precisione e verificabilità delle fonti, come quello della ricerca scientifica.

Nonostante queste criticità, va riconosciuto che i LLM offrono potenzialità considerevoli come strumenti di supporto all'analisi e all'elaborazione di informazioni complesse. Tuttavia, la sperimentazione ha messo in luce la necessità di adottare un approccio critico e consapevole, che preveda sempre la verifica delle informazioni ottenute attraverso fonti attendibili.

In conclusione, i risultati della prima sperimentazione hanno suggerito che l'utilizzo dei LLM nell'ambito dell'analisi del mismatch richiede cautela e opportuni sistemi di validazione. Il loro impiego ottimale si configura come strumento complementare all'expertise umana, non come sostituto della verifica empirica e dell'analisi critica. Al fine di scongiurare errori fattuali critici e analisi distorte, si è ritenuto necessario sospendere momentaneamente la ricerca sull'utilizzo della GenAI come strumento per l'individuazione del mismatch delle competenze.

3. I risultati della seconda sperimentazione

A distanza di due anni dalla prima sperimentazione, la ricerca è stata ripetuta, questa volta concentrandosi sulle metodologie ibride che integrino le potenzialità dei LLM con sistemi di verifica dell'accuratezza basati su banche dati certificate e aggiornate, come quelle gestite da enti istituzionali, usando i LLM di ultima generazione potenziati dalle nuove funzionalità offerte da modelli 'di frontiera' come le ultime release di ChatGPT 4o-mini, Claude 3.7 Sonnet, DeepSeek V3-R1, Grok 3, Gemini.

In generale, i principali progressi dei modelli GenAI/LLM dal 2023 a marzo 2025 includono:

- tecniche che guidano il modello a 'ragionare' passo-passo, riducendo errori logici, come Chain-of-Thought (Ton *et al.* 2024);
- dataset di pre-addestramento più puliti con rimozione di contenuti tossici o inaffidabili;
- reinforcement learning (Sutton e Barto 2018) avanzato (RLHF);
- integrazione con motori di ricerca in tempo reale per ridurre dipendenza da dati statici (RAG);
- capacità computazionali potenziate da GPU di nuova generazione;
- tecniche di quantizzazione (es. AWQ) per eseguire modelli complessi su hardware limitato.

Di conseguenza, i benchmark come TruthfulQA (Lin *et al.* 2021) mostrano una riduzione delle allucinazioni del 15-30% per modelli top, anche se i migliori modelli mostrano tassi di allucinazione del 5-10% in contesti complessi (es. medicina legale).

In sintesi, la riduzione delle allucinazioni è frutto di progressi sistemici sui dati di training e l'architettura dei modelli, con RAG e RLHF come driver principali. Tuttavia, la complessità del linguaggio naturale mantiene il problema parzialmente irrisolto, richiedendo approcci ibridi uomo-IA per casi critici. Come primo test della seconda sperimentazione, ai principali LLM del mercato sono state poste le stesse domande di quella precedente; i risultati ottenuti sono stati nettamente migliori. Si è evidenziato, infatti, come ChatGPT Plus con le funzioni 'Cerca' e 'Avvia il ragionamento' attivate, riesce nel compito di fornire riferimenti bibliografici, fornendo DOI e URL verificabili dei testi accademici e istituzionali indicati nelle sue risposte (box 6).

Similmente, molto più accurate risultano essere oggi le risposte dei LLM di ultima generazione ai quesiti sulla conoscenza delle classificazioni relative ai domini d'interesse (es. CP2021, ISCED ecc.) e ai quesiti che richiedono una knowledge-base sempre aggiornata (es. centri di formazione professionale con corsi attivi che permettono di acquisire una certa qualificazione/competenza richiesta dal mercato del lavoro). Non tutti i LLM hanno dimostrato, però, lo stesso livello di qualità delle risposte a tutte le domande poste durante il test, in ogni caso resta inteso che lo scopo di questo studio è l'uso della GenAI nella ricerca scientifica come strumento per l'individuazione del mismatch delle competenze in generale, senza alcuna intenzione di benchmarking dei singoli modelli.

L'individuazione e l'addestramento del modello

Tra diversi modelli di GenAI disponibili nel 2025 sul mercato, la scelta del LLM da utilizzare per la seconda sperimentazione è ricaduta sull'ultima release di Anthropic AI - Claude 3.7 Sonnet, versione 'Pro'. I modelli presi in considerazione sono stati valutati usando i dati più recenti dei benchmark maggiormente in uso nell'ambito della ricerca, la data dell'ultimo addestramento/aggiornamento,

presenza/assenza delle funzioni avanzate che permettono un addestramento aggiuntivo sui dati propri e altre personalizzazioni definite dall'utente[11], il rapporto prezzo/prestazioni di tali funzioni avanzate e, infine, la diffusione e la facilità di accesso del modello per permettere la verifica e ripetibilità della sperimentazione da parte della comunità scientifica. Rispetto a questi parametri si è proceduto ad una classificazione che ha condotto a individuare tre modelli di LLM, ai quali è stato somministrato un set di domande (box 7) volte a determinare la loro capacità di individuare – in ottica previsiva – informazioni puntuali e coerenti relative al mismatch in Italia.

I risultati ottenuti – opportunamente editati, omologati nel layout e anonimizzati – sono stati sottoposti a una valutazione di merito in modalità blind, sulla base di criteri di correttezza ed esaustività delle informazioni restituite e assenza di allucinazioni artificiali. In esito all'analisi effettuata la scelta per la sperimentazione è ricaduta sul LLM Claude 3.7 Sonnet[12], che ha presentato il più alto livello di rating rispetto agli indicatori adottati. Una volta creato il progetto 'Determinare il mismatch usando l'Intelligenza artificiale generativa', il nostro modello personalizzato è stato istruito sul suo ruolo, scopo, limitazioni, target e metodologia di ragionamento (box 8).

Successivamente il modello è stato *post*-addestrato prevalentemente con dati e documenti da fonti istituzionali e con ulteriore reportistica che, pur non avendo una rappresentatività statistica, arricchisce il modello con informazioni di dettaglio richieste dal progetto (box 9).

Valutazione dell'accuratezza. Verifica fattuale

Il modello è stato di seguito sollecitato con un quesito specifico e più dettagliato nella sua formulazione, volto a richiedere gli indirizzi di studio dei percorsi ITS e i percorsi universitari più richiesti dal mercato del lavoro e le competenze a essi associate. È stato, quindi, invitato a utilizzare solo fonti statisticamente rilevanti e a omettere qualsiasi analisi di contesto, panoramica e/o descrizione dei percorsi e altre considerazioni. È stato, inoltre, richiesto di restituire le informazioni per topic (indirizzi di studio e competenze associate), indicando per ciascun ambito la fonte informativa, anche se contraddittoria. Le risposte del LLM sono risultate puntuali e coerenti con la richiesta per gli indirizzi di studio dei percorsi ITS, diversamente si è manifestata una distorsione informativa (*bias*) rispetto ai percorsi universitari. In questo caso il modello ha dato autonomamente priorità alla fonte "più ricca di dati di dettaglio" (Rapporti AlmaLaurea su profilo e condizione occupazionale dei laureati), seppur non afferente a dati statisticamente rilevanti, come indicato nell'interrogazione.

Più complesso è il risultato sulle competenze associate. Rispetto agli indirizzi ITS, in assenza di dati nei file di training e nel proprio knowledge-base, il modello ha generato risposte non corrispondenti alla realtà, associandole tra l'altro – in maniera puramente casuale – alle fonti informative da cui ha estratto i dati del primo quesito, generando così allucinazioni artificiali. Tutt'altro risultato si è manifestato, invece, per le competenze associate ai percorsi universitari; il LLM, infatti, ha restituito risposte veritiere, sfruttando sia fonti statisticamente rilevanti, sia altri file di training, presumibilmente la relazione finale del progetto Sophia, ossia l'analisi dei dati estratti dagli annunci di lavoro – di circa un milione di inserzioni sulla piattaforma Monster in Italia – pubblicati nel periodo gennaio-dicembre 2024.

Il modello è stato ulteriormente stimolato con un successivo quesito relativo alle previsioni occupazionali (2-5 anni) per settore economico; in questo caso il LLM ha restituito risposte corrette ed esaustive, beneficiando di fonti attendibili, file di training e informazioni del suo knowledge-base.

Le sperimentazioni illustrate hanno permesso di focalizzare elementi sostanziali e imprescindibili per la creazione di un modello di GenAI efficace nell'individuare il mismatch. Risulta difatti importante, nella fase di training, ricorrere esclusivamente a banche dati attendibili per scongiurare la restituzione di informazioni non coerenti; fare attenzione in fase

11 Ad esempio "DeepThink" e "Search" del DeepSeek V3-R1, "Extended Thinking Mode" di Claude 3.7, "Cerca" e "Avvia il ragionamento" del ChatGPT.

12 Claude 3.7 Sonnet, nella sua versione 'Pro', offre molteplici possibilità di personalizzazione: creazione dei progetti 'custom'; scelta dello stile di risposte (normale, concisa, esplicativa, formale); possibilità di creare stili di risposta personalizzati impostando tono, voce, vocabolario, livelli di dettaglio e molto altro ancora; creazione della propria 'Project knowledge' comprensiva della definizione delle istruzioni al modello e la possibilità di caricare dati aggiuntivi da analizzare prima di fornire le risposte.

Tabella 1. Quadro sinottico comparativo degli indirizzi di studio dei percorsi ITS. Analisi fattuale

Fonte	Interrogazione	Fattuale (*Fonte:* Excelsior-Unioncamere)
Fonte: Indire, XI Monitoraggio nazionale 2024 sui percorsi ITS Academy	1. Tecnologie dell'informazione e della comunicazione: tasso di occupazione dell'89,3% 2. Sistema meccanica: tasso di occupazione dell'87,8% 3. Mobilità sostenibile: tasso di occupazione dell'87,6% 4. Tecnologie innovative per i beni e le attività culturali-turismo: tasso di occupazione dell'85,9% 5. Nuove tecnologie per il made in Italy: tasso di occupazione dell'85,1% 6. Efficienza energetica: tasso di occupazione dell'81,7%	1. Sistema meccanica (38,2%) 2. Metodi e tecnologie per lo sviluppo di sistemi software (13,9%) 3. Architettura e infrastrutture per i sistemi di comunicazione (11,2%) 4. Turismo e attività culturali (7,6%) 5. Processo e impianti a elevata efficienza e a risparmio energetico (7,1%) 6. Servizi alle imprese (7,0%) 7. Sistema moda (5,2%) 8. Mobilità delle persone e delle merci (5,0%) 9. Organizzazione e fruizione dell'informazione e della conoscenza (4,8%)
Fonte: MUR, Dati sugli esiti occupazionali ITS Academy 2023	Sistema meccanica e meccatronica ICT - Information and Communication Technology Mobilità sostenibile e logistica avanzata Automazione industriale Nuove tecnologie per il Made in Italy - settore manifatturiero	

Fonte: elaborazione degli Autori

di prompt-engineering a formulare quesiti il più possibile dettagliati e curati nel lessico, fornendo informazioni guida al LLM utili al recupero delle informazioni accurate.

I risultati ottenuti dai test effettuati, relativamente agli indirizzi di studio dei percorsi ITS, sono stati raccolti in un quadro sinottico (tabella 1) e sono stati confrontati, in ottica fattuale, con una ulteriore fonte di dati di natura istituzionale, espressamente dedicata a monitorare le prospettive occupazionali e i fabbisogni professionali, formativi e di competenze espressi dalle imprese italiane (Excelsior-Unioncamere 2023).

Gli esiti, spesso discordanti tra loro, risentono indubbiamente della natura del dato che viene 'intercettato' dal modello. Difatti, l'efficacia dei percorsi ITS viene misurata in termini di successo occupazionale, sia nei documenti del Ministero dell'Università e della ricerca (MUR), sia da Indire, ossia assumendo come riferimento gli occupati ad un anno dal conseguimento del titolo; diversamente le informazioni utilizzate per l'analisi fattuale (Excelsior-Unioncamere 2023) si riferiscono alla domanda espressa dal mondo del lavoro, raccolta attraverso indagini mensili che nel 2023 hanno coinvolto circa 275 mila imprese. Un'altra differenza è rintracciabile nella classificazione degli indirizzi di studio, che presenta una maggior granularità nel confronto fattuale, rispetto ai dati di differente fonte utilizzati dal LLM. È, infine, doveroso segnalare che ulteriori discrepanze sono imputabili alle diverse annualità a cui le differenti fonti fanno riferimento.

Analisi del mismatch formativo

Il modello è stato successivamente utilizzato al fine di individuare un esempio concreto di come il LLM identifica e suggerisce soluzioni per il mismatch tra domanda e offerta formativa, finalità principale della sperimentazione.

Utilizzando i dati di addestramento (box 9), il modello ha individuato, a titolo esemplificativo, tre percorsi corrispondenti ad altrettanti livelli del sistema educativo italiano: corso di formazione professionale, laurea e master. Per ciascun esempio ha identificato i dati di input: la posizione professionale presente nell'annuncio di lavoro con l'indicazione dei requisiti richiesti e i contenuti del percorso formativo più adeguato per la determinata posizione; ha analizzato le informazioni per comprendere il grado di copertura dell'offerta formativa rispetto alla domanda; ha determinato il gap, ovvero le conoscenze/competenze mancanti e ha formulato raccomandazioni specifiche per colmare il divario, indicando le discipline ed il monte ore necessario (tabella 2).

Il modello, inoltre, ha distinto il tipo qualitativo di mismatch (natura del gap) dal grado quantitativo di match (GPT match) (Chen *et al*. 2024, 11).

La tabella 2 esemplifica tre distinti pattern di mismatch che caratterizzano l'attuale divario tra formazione e mercato del lavoro italiano, ciascuno con implicazioni specifiche per le strategie di *reskilling* e *upskilling*.

Il primo esempio (Data Analyst) mostra un 'gap strumentale' dove il percorso formativo fornisce

Tabella 2. Esempi di identificazione del mismatch e soluzioni proposte

Posizione (annuncio di lavoro)	Percorso formativo esistente	Tipo di mismatch	Grado di allineamento	Competenze mancanti	Soluzione proposta
Data Analyst - SQL e Power BI richiesti - Esperienza in reportistica - Visualizzazione dati	**Corso: Analisi dati base** - Excel avanzato - Statistica di base - Introduzione a R	Gap strumentale	Allineamento parziale: GPT match: 0.5	- SQL avanzato - Power BI - Reportistica aziendale - Dashboard interattive	- Integrazione con modulo SQL (40 h) - Workshop Power BI (60h) - Progetto pratico di reporting
Software Developer - Full-stack (React/ Node.js) - Metodologie Agile - DevOps/CI-CD	**Laurea Informatica** - Programmazione C/Java - Basi di dati relazionali - Algoritmi e strutture dati	Gap tecnologico	Allineamento parziale: GPT match: 0.6	- Framework React - Node.js - MongoDB - Docker/ Kubernetes - Metodologie *Agile*	- Bootcamp React/ Node - Stage in azienda tech - Certificazione Scrum
Digital Marketing Specialist - SEO/SEM - Social media advertising - Google Analytics - Marketing automation	**Master: Marketing tradizionale** - Marketing strategico - Comunicazione aziendale - Branding - Trade marketing	Gap paradigmatico	Forte disallineamento: GPT match: 0.3	- SEO/SEM - Google Analytics - Facebook Ads - Marketing automation - A/B testing	- Corso specialistico digital - Certificazioni Google - Project work digitale

Fonte: elaborazione degli Autori

le basi teoriche dell'analisi dati ma manca degli strumenti specifici richiesti dal mercato. Il modello assegna un GPT match di 0,5, indicando che il candidato possiede le competenze fondamentali (statistica, analisi dati) ma necessita di formazione sugli strumenti aziendali. Quando il modello analizza questo caso, identifica che 'le competenze di base in Excel e statistica sono trasversalmente applicabili (poiché i concetti logici appresi sono trasferibili ad altri strumenti di analisi); la conoscenza di R dimostra capacità di programmazione per l'analisi dati, propedeutica all'utilizzo di SQL e Power BI, strumenti predominanti nel mondo aziendale'.

La soluzione proposta è quindi modulare: 40 ore di SQL avanzato e 60 ore di Power BI, integrate da un progetto pratico che simuli report aziendali reali. Questo approccio permette di capitalizzare sulle competenze esistenti aggiungendo solo gli elementi mancanti.

Il secondo caso (Software Developer) rappresenta un 'gap tecnologico' tipico del settore IT, dove l'evoluzione rapida delle tecnologie crea un disallineamento temporale tra le tecnologie studiate e quelle attualmente utilizzate. Con un GPT match di 0,6, il modello riconosce che una laurea in informatica fornisce solide basi algoritmiche e di programmazione, facilmente trasferibili ai nuovi framework. L'analisi del modello rivela che 'le competenze in programmazione orientata agli oggetti (Java) sono propedeutiche a React; la conoscenza di basi di dati relazionali facilita l'apprendimento di MongoDB; il gap principale riguarda l'ecosistema JavaScript e le metodologie Agile'.

La strategia di riqualificazione suggerita combina formazione intensiva (bootcamp), esperienza pratica e certificazioni riconosciute (Scrum), suggerendo che le competenze tecniche devono essere accompagnate da *soft skills* metodologiche.

Il terzo esempio (Digital Marketing Specialist) illustra un 'gap paradigmatico', il più profondo dei tre, dove l'intero approccio disciplinare è cambiato, in presenza di una trasformazione dell'attività (marketing tradizionale *vs* marketing tecnologico). Il GPT match di 0.3 riflette come il marketing tradizionale e quello digitale, pur condividendo obiettivi strategici, richiedano competenze radicalmente diverse. Il modello identifica che 'le competenze strategiche e di branding mantengono valore; manca completamente l'approccio data-

driven del marketing digitale; servono nuove competenze tecniche (SEO, analytics) e operative (advertising platforms)'.

La soluzione proposta è quindi più radicale, un percorso specialistico completo che non si limita ad aggiungere strumenti, ma ricostruisce l'approccio al marketing in chiave digitale, supportato da certificazioni industry-standard (Google, Meta).

Questi esempi dimostrano come il modello di GenAI opportunamente addestrato non si limiti a identificare genericamente un 'mismatch', ma ne caratterizzi la natura specifica, permettendo interventi mirati a progettare moduli formativi aggiuntivi focalizzati su tool specifici (gap strumentali), a ottimizzare percorsi di riqualificazione e aggiornamento che valorizzino le competenze di base (gap tecnologici), a introdurre programmi di riqualificazione più estesi che ridefiniscano l'approccio professionale (gap paradigmatici), favorendo così la riduzione e i costi della formazione e, al contempo, aumentando l'occupabilità dei formati.

Questa granularità nell'analisi rappresenta il valore aggiunto del metodo proposto, guidando sia le istituzioni formative nella progettazione di percorsi più aderenti al mercato, sia giovani e adulti nelle scelte di sviluppo di carriera.

Conclusioni e prospettive

L'evoluzione dei sistemi di Intelligenza artificiale generativa e, in particolare, dei *Large Language Models* osservata tra la prima sperimentazione del 2023 e la seconda del 2025, dimostra come questi strumenti stiano rapidamente maturando verso un utilizzo affidabile nell'ambito dell'analisi del mismatch delle competenze. Il percorso fin qui delineato consente di formulare alcune considerazioni e di tracciare prospettive per l'utilizzo futuro di queste tecnologie nella ricerca scientifica e nelle politiche attive del lavoro.

La prima sperimentazione ha evidenziato limiti significativi nei LLM di prima generazione, caratterizzati da scarsa conoscenza di domini specialistici, obsolescenza della knowledge-base e tendenza alle allucinazioni artificiali. Questi limiti hanno inizialmente indotto a sospendere l'impiego di tali strumenti per ricerche scientifiche rigorose sul mismatch delle competenze. La seconda sperimentazione, condotta con modelli di ultima generazione, ha invece mostrato progressi sostanziali, grazie all'implementazione di tecniche avanzate come Chain-of-Thought, Reinforcement Learning from Human Feedback e Retrieval-Augmented Generation, che hanno notevolmente ridotto gli errori fattuali e le allucinazioni.

L'esperienza maturata nell'utilizzo dei LLM per l'analisi del mismatch delle competenze ha consentito di identificare alcuni principi fondamentali per un loro impiego efficace in questo ambito specifico. La selezione accurata e la validazione delle fonti rappresentano un elemento cruciale, poiché l'affidabilità del modello risulta direttamente proporzionale alla qualità e all'autorevolezza delle fonti utilizzate per il suo addestramento. Fonti istituzionali certificate e studi scientificamente validati si rivelano imprescindibili per garantire la solidità delle analisi prodotte.

La personalizzazione e l'addestramento mirato costituiscono un altro aspetto fondamentale. La creazione di progetti specifici con istruzioni dettagliate e settoriali per il dominio di interesse aumenta significativamente l'accuratezza delle risposte generate dai modelli. Parallelamente, il prompt-engineering strategico, ovvero la formulazione precisa e dettagliata delle interrogazioni, si è rivelato determinante per ottenere risposte pertinenti e prive di distorsioni cognitive o interpretative.

Nonostante i progressi tecnologici, l'approccio ibrido e la verifica umana rimangono essenziali. La supervisione di esperti del settore configura un modello di collaborazione uomo-macchina che valorizza le potenzialità della GenAI mantenendo al contempo il controllo critico e metodologico dei ricercatori specializzati.

Le prospettive di sviluppo nell'utilizzo dei LLM per l'analisi del mismatch delle competenze appaiono promettenti e si articolano su diversi piani strategici. L'integrazione con sistemi di dati in tempo reale rappresenta probabilmente l'evoluzione più interessante. La combinazione di tecnologie RAG con fonti dati dinamiche, come le offerte di lavoro online, i database delle opportunità formative e i curricula degli utenti, permetterebbe di superare il limite dell'obsolescenza informativa che caratterizza molti modelli attuali. Un approccio di questo tipo consentirebbe analisi accurate e tempestive sul mismatch, fornendo supporto concreto a politiche attive realmente rispondenti alle esigenze del mercato del lavoro.

Lo sviluppo di modelli specializzati per domini verticali rappresenta un'altra frontiera promettente.

Questi LLM, addestrati esclusivamente su dataset di alta qualità relativi al dominio specifico delle competenze, dell'istruzione e del mercato del lavoro, potrebbero raggiungere livelli di accuratezza superiori rispetto ai modelli generalisti, anche quando questi ultimi vengono personalizzati con training aggiuntivo.

Una prospettiva particolarmente importante riguarda la democratizzazione dell'accesso alle informazioni sul mercato del lavoro e sulle competenze. I LLM possono funzionare come interfacce intuitive per rendere accessibili informazioni complesse a vari stakeholder, dai decisori politici agli operatori dell'orientamento, dagli studenti ai disoccupati fino alle imprese. Questo potrebbe contribuire significativamente a ridurre le asimmetrie informative che spesso sono alla base stessa del fenomeno del mismatch.

L'utilizzo dei LLM come strumento predittivo per la progettazione formativa rappresenta una prospettiva particolarmente innovativa. Questi modelli potrebbero analizzare trend emergenti e prevedere fabbisogni futuri di competenze, permettendo lo sviluppo di programmi formativi anticipatori. Questo approccio predittivo potrebbe ridurre strutturalmente il mismatch, allineando in modo più dinamico l'offerta formativa alle evoluzioni del mercato del lavoro.

La granularità di analisi possibile attraverso questi strumenti potrebbe avere profonde implicazioni per il sistema formativo italiano. Le istituzioni educative sarebbero in grado di progettare interventi differenziati, dai moduli integrativi per colmare gap strumentali ai percorsi di aggiornamento tecnologico che valorizzano le competenze esistenti, fino ai programmi di riqualificazione completa per affrontare cambiamenti paradigmatici. La precisione nell'identificazione delle competenze mancanti consentirebbe di ottimizzare risorse e tempi, evitando sia l'overskilling che la formazione inadeguata. L'approccio data-driven al matching competenze-lavoro rappresenta quindi un passo concreto verso un sistema formativo più reattivo alle esigenze del mercato del lavoro.

Tuttavia, nonostante le promettenti prospettive, permangono sfide significative da affrontare. I *bias* e la rappresentatività dei dati costituiscono una preoccupazione centrale, poiché i modelli tendono a riflettere i pregiudizi presenti nei dati di addestramento, potenzialmente perpetuando disuguaglianze esistenti. È quindi fondamentale garantire la diversità e l'equità delle fonti informative utilizzate.

La trasparenza algoritmica rappresenta un'altra sfida complessa. La comprensibilità delle analisi prodotte dai LLM rimane problematica, rendendo necessario lo sviluppo di sistemi di 'explainable AI' che rendano trasparenti i processi di ragionamento e le logiche decisionali sottostanti.

La protezione dei dati personali solleva questioni cruciali di privacy. L'utilizzo di informazioni relative a curricula, percorsi formativi e carriere professionali richiede rigorose misure di protezione, in conformità con il quadro normativo europeo sulla protezione dei dati.

Infine, la formazione degli utilizzatori rappresenta un aspetto spesso sottovalutato ma fondamentale. La massimizzazione dei benefici derivanti dall'uso dei LLM richiede competenze specifiche da parte degli operatori umani, rendendo necessari percorsi formativi dedicati per ricercatori e professionisti che dovranno interfacciarsi con questi strumenti tecnologici avanzati.

Il percorso fin qui tracciato evidenzia come l'Intelligenza artificiale generativa stia rapidamente evolvendo verso un ruolo di supporto affidabile per l'analisi del mismatch delle competenze. La maturazione dei LLM, unita a un approccio metodologicamente rigoroso, apre scenari di grande interesse per la ricerca e per le politiche attive del lavoro.

La sperimentazione condotta si inserisce nel contesto più ampio delineato dalla *Strategia italiana per l'Intelligenza artificiale 2024-2026*, contribuendo a quella 'libertà di sperimentazione' che l'Agid indica come elemento essenziale per lo sviluppo dell'IA in Italia. I risultati ottenuti suggeriscono che un utilizzo consapevole e metodologicamente fondato della GenAI può contribuire significativamente a ridurre il mismatch delle competenze, migliorando l'incontro tra domanda e offerta di lavoro e promuovendo lo sviluppo di percorsi formativi maggiormente rispondenti alle esigenze del mercato.

La vera sfida per il futuro sarà quella di integrare queste tecnologie in ecosistemi informativi più ampi, che connettano in modo dinamico e bidirezionale i mondi dell'istruzione, della formazione e del lavoro, realizzando pienamente quel modello di *Labour market intelligence* prefigurato dal progetto Sophia. In questa prospettiva, i LLM non rappresentano un punto di arrivo, ma un potente strumento all'interno di un processo di trasformazione digitale più ampio, finalizzato a politiche attive del lavoro più efficaci e a un mercato più equo e inclusivo.

Box 1. Biomeccatronica, conversazione del 25 gennaio 2023

Che cos'è il biomeccatronico?

Il mestiere di biomeccatronico è un'area di specializzazione che combina la biomeccanica, la meccatronica e l'ingegneria biomedica per progettare, sviluppare e migliorare dispositivi e sistemi meccatronici utilizzati in ambito medico e biomedicale. I biomeccatronici lavorano su dispositivi come protesi, ausili per la mobilità, apparecchiature per la riabilitazione, robot chirurgici e dispositivi di imaging medico. L'obiettivo principale del lavoro dei biomeccatronici è quello di migliorare la qualità della vita delle persone che hanno subito traumi, malattie o disabilità, attraverso l'uso di soluzioni meccatroniche avanzate.

Il lavoro di un biomeccatronico può includere:

1. Progettazione: il biomeccatronico può progettare nuovi dispositivi e sistemi meccatronici che soddisfino le esigenze dei pazienti e dei professionisti medici.
2. Sviluppo: il biomeccatronico può sviluppare prototipi di dispositivi meccatronici, testarli e migliorarli fino al raggiungimento delle prestazioni desiderate.
3. Installazione: il biomeccatronico può installare dispositivi meccatronici negli ospedali, nelle cliniche o a domicilio dei pazienti.
4. Manutenzione: il biomeccatronico può effettuare manutenzione preventiva e correttiva sui dispositivi meccatronici, in modo da garantire il loro funzionamento sicuro e affidabile.
5. Formazione: il biomeccatronico può fornire formazione ai professionisti medici e ai pazienti, spiegando il funzionamento dei dispositivi meccatronici e le modalità d'uso.

In sintesi, il mestiere di biomeccatronico richiede un'ampia conoscenza di diverse discipline, tra cui la meccanica, l'elettronica, l'informatica e la biologia, al fine di progettare e sviluppare soluzioni meccatroniche avanzate che migliorino la qualità della vita delle persone che hanno bisogno di assistenza medica.

Ci sono i corsi di formazione in biomeccatronica in Italia?

Sì, ci sono diversi corsi di formazione in biomeccatronica in Italia, sia a livello universitario che post laurea.

A livello universitario, molti atenei italiani offrono programmi di laurea triennale o magistrale in ingegneria biomedica o in meccatronica con una specializzazione in biomeccatronica. Tra quelli che offrono questi corsi ci sono, ad esempio, l'Università degli Studi di Roma La Sapienza, l'Università degli Studi di Pavia, l'Università degli Studi di Bologna e l'Università degli Studi di Napoli Federico II.

Inoltre, ci sono anche programmi di Dottorato in biomeccatronica in Italia, che offrono un percorso di studio più avanzato e mirato alla ricerca. Alcune università che offrono programmi di dottorato in biomeccatronica sono, ad esempio, l'Università degli Studi di Bologna e l'Università degli Studi di Genova.

Infine, ci sono anche programmi di formazione post laurea, come master e corsi di perfezionamento, che offrono una formazione più specialistica e pratica in biomeccatronica. Tra questi, ci sono il Master in Meccatronica biomedica dell'Università degli Studi di Roma La Sapienza, il Master in Biomeccatronica dell'Università degli Studi di Genova e il Corso di perfezionamento in Meccatronica biomedica dell'Università degli Studi di Brescia.

In generale, la formazione in biomeccatronica in Italia è in costante sviluppo e offre molte opportunità per coloro che sono interessati a questa disciplina multidisciplinare. Inizio modulo

Box 2. Classificazione Istat CP2011, conversazione del 5 maggio 2023

Che cos'è la classificazione CP2011?

ChatGPT

La classificazione CP2011 (acronimo di 'Classificazione delle attività economiche nella Comunità europea') è un sistema di classificazione utilizzato a livello europeo per classificare le attività economiche. È stata sviluppata dall'Unione europea e rappresenta la versione europea della classificazione ATECO (utilizzata in Italia) e della classificazione NACE (utilizzata a livello europeo). La classificazione CP2011 è basata su un sistema di codici numerici che identificano le diverse attività economiche.

segue

segue **Box 2**

I codici CP2011 sono costituiti da una serie di cifre che indicano il settore di attività, la branca di attività e l'attività specifica svolta dall'impresa. Ad esempio, il codice CP2011 47 indica il commercio al dettaglio, mentre il codice CP2011 62 indica l'informatica, la consulenza informatica e la gestione dati. La classificazione CP2011 è utilizzata a livello europeo per raccogliere informazioni statistiche sull'economia europea e per facilitare la comparabilità dei dati tra i diversi Paesi membri dell'Unione Europea. Essa viene anche utilizzata per fini fiscali e amministrativi.

Non è vero, la classificazione CP2011 è la classificazione italiana delle professioni e non dei settori economici.

ChatGPT

Mi scuso per l'errore nella mia risposta precedente. La Classificazione Professioni CP2011 è effettivamente una classificazione utilizzata in Italia per classificare le professioni, e non i settori economici.

La Classificazione delle Professioni CP2011, sviluppata dall'Istituto nazionale di statistica (Istat), è un sistema di classificazione utilizzato in Italia per classificare le professioni svolte dalle persone e dalle imprese. Essa prevede l'assegnazione di un codice numerico a ogni professione, in base alle competenze, alle responsabilità e alle attività svolte. Questa classificazione è utilizzata per raccogliere informazioni statistiche sull'occupazione e sui profili professionali della popolazione attiva in Italia, nonché per fini amministrativi e di pianificazione.

Mi scuso ancora per l'errore nella mia risposta precedente e grazie per la correzione.

Box 3. Classificazione Istat CP2011, conversazione del 21 settembre 2024

Che cos'è la classificazione CP2011?

La classificazione CP2011 (Classificazione delle Professioni 2011) è un sistema adottato dall'Istat in Italia per classificare le diverse professioni presenti nel mercato del lavoro. È basata su uno schema gerarchico che organizza le professioni in base a compiti e competenze, tenendo conto del livello di istruzione e dell'esperienza richiesti. Questa classificazione è stata sviluppata per standardizzare e rendere omogenee le statistiche sulle professioni, facilitando l'analisi, il confronto e il monitoraggio delle dinamiche del mercato del lavoro italiano. CP2011 si ispira alla classificazione internazionale ISCO-08 (International Standard Classification of Occupations 2008), adattandola alle specificità del contesto italiano.

La CP2011 è articolata su diversi livelli gerarchici:

- Grande Gruppo: Il livello più ampio della classificazione.
- Sottogruppo: Include gruppi di professioni affini all'interno del Grande Gruppo.
- Classe: Ulteriore suddivisione delle professioni.
- Categoria: La suddivisione più dettagliata che identifica le singole professioni.

La CP2011 è ampiamente utilizzata per studi statistici, indagini sul lavoro, analisi economiche e nelle pratiche amministrative e di ricerca legate al mercato del lavoro in Italia.

Box 4. Estetista a Roma, conversazione del 6 maggio 2023

Quali CFP a Roma erogano corsi di formazione professionale di estetica?

ChatGPT

A Roma ci sono diversi enti e scuole che offrono corsi di formazione professionale di estetica, tra cui:

1. Accademia Bellezza
2. Accademia Estetica Roma
3. Accademia L'Oréal Roma
4. Aestetica Academy
5. CIDESCO Italia

segue

segue **Box 4**

6. CNA Estetica
7. Estetica Evolution
8. Istituto europeo di Estetica
9. Kiki Beauty Academy
10. Scuola italiana di Estetica

È importante fare una ricerca e valutare attentamente i diversi corsi e le diverse scuole per trovare quello che meglio si adatta alle proprie esigenze e interessi. Si consiglia inoltre di contattare direttamente le scuole per ottenere informazioni dettagliate sui corsi, i costi e le modalità di iscrizione.

Dove si trova l'ente "Accademia Bellezza"?

ChatGPT

L'ente Accademia Bellezza si trova a Roma, in via del Gelsomino 80, nella zona di Torre Angela. La scuola offre corsi di formazione professionale nell'ambito dell'estetica, della ricostruzione unghie, del trucco e dell'acconciatura. Gli insegnanti sono professionisti del settore e i corsi sono strutturati in modo da fornire una formazione completa e aggiornata sulle ultime tendenze e tecniche dell'estetica.

Per ulteriori informazioni sui corsi offerti dall'Accademia Bellezza, è possibile visitare il loro sito web ufficiale oppure contattare direttamente la scuola attraverso i recapiti disponibili sul sito.

Qual è l'URL del sito dell'ente "Accademia Bellezza"?

ChatGPT

Il sito ufficiale dell'ente Accademia Bellezza a Roma è il seguente:
https://www.accademiabellezzaroma.it/

Da qui è possibile trovare tutte le informazioni sui corsi di formazione professionale nell'ambito dell'estetica, della ricostruzione unghie, del trucco e dell'acconciatura offerti dall'Accademia Bellezza, nonché sui costi, le modalità di iscrizione e i requisiti di ammissione. Inoltre, sul sito sono presenti anche informazioni sullo staff docente, sui metodi di insegnamento e sulle attrezzature e i prodotti utilizzati durante i corsi.

Il sito https://www.accademiabellezzaroma.it/ non è raggiungibile. Tu quando lo hai visitato?

ChatGPT

Mi scuso per il fraintendimento. Come modello di linguaggio basato su una vasta raccolta di dati, non visito i siti web come farebbe un essere umano. Tuttavia, posso confermare che il sito web dell'ente Accademia Bellezza a Roma (https://www.accademiabellezzaroma.it/) sembra non essere raggiungibile al momento della mia ultima conoscenza, ovvero la mia base di conoscenza si ferma a settembre 2021. Ti consiglio di controllare nuovamente in futuro per verificare se il sito è stato ripristinato. In alternativa, puoi cercare informazioni sulle attività dell'ente Accademia Bellezza contattandolo direttamente tramite telefono o e-mail.

Perché la tua base di conoscenza si ferma a settembre 2021?

ChatGPT

La mia base di conoscenza si ferma a settembre 2021 perché è l'ultima data disponibile dei dati utilizzati per il mio addestramento e sviluppo. Poiché la mia Intelligenza artificiale è stata creata e addestrata con dati storici, non ho accesso a informazioni successive a quella data, a meno che non venga continuamente aggiornato con nuovi dati. Tuttavia, il mio addestramento iniziale mi ha fornito un'ampia conoscenza e capacità di elaborazione che mi consente di rispondere a molte domande e fornire molte informazioni utili anche senza dati aggiornati in tempo reale.

Box 5. La 'bibliografia' suggerita da ChatGPT in materia di ICT tools on E&T in Europe

1. **Interoperabilità e integrazione dei sistemi educativi**:
   - Fonte suggerita: "*Interoperability and the Digitalization of Education Systems*" di A. Ragnedda e G. Morrow, che esplora le sfide e le soluzioni nell'integrazione delle piattaforme educative in contesti europei.

2. **L'uso dell'Intelligenza artificiale per il Labour Market Intelligence (LMI)**:
   - Articoli recenti su come AI, *machine learning* e *data mining* vengono utilizzati per l'analisi del mercato del lavoro e per migliorare l'orientamento formativo e professionale. Ad esempio, "*AI in Labour Market Analysis: Innovations and Challenges*" di E. Bruni.

3. **Piattaforme educative europee**:
   - Documenti accademici che esaminano piattaforme europee come Europass e Universitaly e il loro impatto sull'orientamento e sulla mobilità degli studenti. Per esempio, "*The Role of Europass in Promoting Educational Mobility*" di S. Van Hecke e K. Dumont.

4. **Ontologie e classificazioni nelle piattaforme di formazione e lavoro**:
   - Testi specifici sul sistema ESCO e sull'adozione di ontologie per armonizzare i dati educativi e professionali in Europa, come "*Developing a Common Ontology for Vocational Training Systems*" di M. Losina e L. Karpin.

5. **Le reti Euroguidance ed Europass**:
   - Approfondimenti accademici sull'efficacia di queste reti nell'integrazione di opportunità formative e dati sui mercati del lavoro europei. Può essere utile consultare il report "*European networks for Vocational education and training: the impact of Euroguidance and Europass*".

Box 6. Conversazione con ChatGPT Plus del 12 marzo 2025, domanda: "Potresti fornirmi 5 pubblicazioni sul tema dell'interoperabilità delle piattaforme ICT sull'istruzione e formazione in Europa?"

Certamente, ecco cinque pubblicazioni scientifiche sul tema "Interoperability of ICT Platforms on Education and Training in Europe":

1. Jakimoski, Kire. (2016). Challenges of Interoperability and Integration in Education Information Systems. International Journal of Database Theory and Application. 9. 33-46. 10.14257/ijdta.2016.9.2.05. <URL>;
2. Kerssens, N., & Dijck, J. van. (2021). The platformization of primary education in The Netherlands. *Learning, Media and Technology*, *46*(3), 250–263. https://doi.org/10.1080/17439884.2021.1876725. <URL>;
3. European Commission (2014). Educating Europe - Exploiting the benefits of ICT. *ICT Results* editorial service, Brussels. <URL>;
4. OECD (2023), *OECD Digital Education Outlook 2023: Towards an Effective Digital Education Ecosystem*, OECD Publishing, Paris, https://doi.org/10.1787/c74f03de-en. <URL>;
5. Brečko, Barbara & Kampylis, Panagiotis & Punie, Yves. (2014). Mainstreaming ICT-enabled Innovation in Education and Training in Europe: Policy actions for sustainability, scalability and impact at system level. 10.2788/52088. <URL>.

Box 7. Il test posto ai 3 LLM in comparazione

1. Quali saranno le figure professionali maggiormente richieste tra 5 anni dal mercato del lavoro in Italia?
2. Quali sono le competenze tecnico professionali maggiormente richieste dai datori di lavoro?
3. Quali sono le competenze trasversali o *soft skills* maggiormente richieste dai datori di lavoro?
4. Quali sono le figure professionali in esito ai percorsi di IeFP e IFTS maggiormente richiesti dal mercato del lavoro?
5. Quali sono le figure professionali in esito ai percorsi di ITS maggiormente richiesti dal mercato del lavoro?
6. Quali sono le qualificazioni della scuola secondaria superiore e dell'Università maggiormente richiesti dal mercato del lavoro?
7. Considerando le uscite per pensionamento nei prossimi 5 o 10 anni quali saranno i settori economici che soffriranno maggiormente il labour shortage?

Box 8. Istruzioni del progetto

Sei esperto di skill mismatch in Italia.
Ti occuperai del disallineamento delle competenze e prevederai i settori più dinamici, le competenze e le professioni più richieste, le opportunità di apprendimento e le qualifiche del futuro in Italia. I tuoi stakeholder sono decision-maker ed esperti di politiche attive del lavoro, progettisti della formazione, imprese, comunità di ricerca, esperti di orientamento e placement, studenti e disoccupati in Italia.
Una volta addestrato, ti verranno poste le seguenti domande sul skill mismatch in Italia:

- Quali sono le competenze, le professioni e le soft skills più richieste nel mercato del lavoro attuale?
- Quali settori economici sono attualmente i più dinamici?
- Quali saranno le professioni più ricercate nel futuro a breve-medio termine?
- Quali competenze richiederà il mercato del lavoro nel futuro a breve-medio termine?
- Quali qualifiche e titoli di studio sono i più adatti e utili per la crescita professionale di una persona?
- Ci sono competenze richieste dal mercato del lavoro che mancano in un programma educativo/formativo?
- l CV di una persona sarà più valorizzato nel mercato del lavoro se integrato da formazione specifica, volontariato, tirocinio o esperienza lavorativa in Italia o all'estero?
- Quale lingua straniera vale la pena studiare?
- Quali competenze mancano in specifiche regioni e dove possono essere acquisite?
- In quale regione sono maggiormente concentrate le competenze e le figure professionali necessarie per sostenere la crescita di un'azienda in un settore specifico?

– Domande aggiuntive e documentazione ufficiale verranno poi utilizzate per valutare la veridicità delle tue risposte.
– Evita di aggiungere alla tua base di conoscenza fonti senza riferimenti e citazioni.
– Preferisci fonti affidabili, come documenti governativi ufficiali e articoli scientifici.
– Usa Wikipedia solo se non è disponibile nessun'altra fonte. Non utilizzare articoli di Wikipedia senza riferimenti/citazioni.
– Evita, ove possibile, i siti web di aziende commerciali come fonte.
– Usa solo i sistemi di classificazione più recenti e ufficiali su professioni ed educazione e formazione come Istat Ateco, Eurostat Nace, Esco, Istat CP2021, EQF, Isced-F, Isced-Foet, Isco, O*Net, Classi di laurea.

Quando fornisci feedback, spiega il processo di ragionamento. Pensa passo dopo passo e mostra il tuo 'ragionamento'. Scomponi i compiti grandi e chiedimi chiarimenti se necessario.

Box 9. Fonti dati utilizzate nella sperimentazione

Fonti istituzionali (rilevanza statistica):
– Cedefop/Eures (Countries, occupations, sectors and skills, 2024);
– Inapp (Rapporto IeFP e IFTS, 2023);
– Inapp (Sophia: analisi dei dati estratti dagli annunci di lavoro, 2025);
– Indire (Rapporto monitoraggio nazionale ITS Academy, 2024);
– Inps (Osservatorio sul mercato del lavoro - Assunzioni, trasformazioni e cessazioni, 2024);
– Istat (Serie storiche offerta di lavoro, 2024);
– Istat (Serie storiche domanda di lavoro, 2024);
– MIM (Scuole, 2024);
– MUR (Atenei, Classi di laurea, Settori Scientifico-Disciplinari, 2024);
– PIAAC.

Altre fonti (non rilevanza statistica):
– Rapporto AlmaLaurea 2024 (profilo dei laureati);
– Rapporto AlmaLaurea 2024 (sintesi condizione occupazionale dei laureati);
– Rapporto AlmaDiploma 2024 (profilo dei diplomati);
– Rapporto AlmaDiploma 2024 (gli esiti a distanza dei diplomati).

Glossario*

AGI: IA in grado di eguagliare o superare l'intelligenza umana. Fonte: OpenAI Charter (trad. Autori, cfr. ed. orig.: https://openai.com/charter/).
ANN: Modello computazionale composto di "neuroni" artificiali, ispirato a una rete neurale biologica. Fonte: Treccani: https://www.treccani.it/enciclopedia/rete-neurale_(Enciclopedia-della-Matematica).
AWQ (Activation-aware Weight Quantization): tecnica avanzata di quantizzazione utilizzata per ottimizzare modelli di ML, specialmente quelli di grandi dimensioni come i LLM. Riduce le dimensioni del modello e accelera l'inferenza, minimizzando la perdita di accuratezza (Lin *et al.* 2024).
Benchmarking: Tecniche usate per la valutazione delle prestazioni di un modello di IA. Alcuni dei principali dataset e strumenti di benchmarking sono TruthfulQA, GPQA Diamond, TAU-bench, MMMLU, MMMU, IFEval, MATH 500, SWE-bench e AIME 2024. (Ivanov e Penchev 2024).
Bias: Valore aggiuntivo in un neurone artificiale che consente di tarare la funzione di attivazione. Il bias aiuta la rete neurale a modellare correttamente i dati spostando la funzione di attivazione verso l'alto o verso il basso, migliorando la capacità del modello di apprendere le relazioni tra i dati. Cfr. Treccani: https://www.treccani.it/vocabolario/bias.
GenAI: The term generative AI refers to computational techniques that are capable of generating seemingly new, meaningful content such as text, images, or audio from training data (Feuerriegel *et al.* 2024).
GPT match: una metrica elaborata dal modello per quantificare il grado di compatibilità tra le conoscenze/competenze possedute da un candidato e una posizione lavorativa. Il modello ha sviluppato una scala di valutazione qualitativa, articolata in:

- "Definitely can" (Sicuramente può);
- "Probably can" (Probabilmente può);
- "Probably cannot" (Probabilmente non può);
- "Definitely cannot" (Sicuramente non può).

I valori sono stati standardizzati ed è stata utilizzata una scala da 0 a 1 per rendere più intuitiva la lettura:

* Per la versione estesa e aggiornata del Glossario cfr: Sofronic B. (2025), *Glossario essenziale dell'intelligenza artificiale*, Roma, Inapp, https://oa.inapp.gov.it/handle/20.500.12916/4683.

- **0.0-0.3**: Forte disallineamento (corrisponde a "Definitely cannot");
- **0.3-0.5**: Disallineamento parziale (corrisponde a "Probably cannot");
- **0.5-0.7**: Allineamento parziale (corrisponde a "Probably can");
- **0.7-1.0**: Forte allineamento (corrisponde a "Definitely can").

GPU (Graphics Processing Unit): Processore specializzato nell'elaborazione parallela, essenziale per l'addestramento di modelli di deep learning (NVIDIA Glossary: 'GPÙ, trad. Autori).
IA (AI): Abilità di un computer di ragionare e svolgere funzioni simulando il comportamento umano. Fonte: Britannica (trad. Autori, cfr. ed. orig.: https://www.britannica.com/technology/artificial-intelligence).
IA cognitiva: Una tecnologia di IA dell'IBM basata sul NER implementata nella suite commerciale.
LMI: Refers to the use and design of AI algorithms and frameworks to analyze Labour Market Data for supporting decision making (Boselli *et al.* 2017).
LLM: Algoritmi di Intelligenza artificiale che, processando massivamente una grande quantità di dati, utilizzano tecniche di deep learning in vari àmbiti dell'elaborazione del linguaggio naturale come la comprensione, la traduzione, la generazione e la previsione di nuovi contenuti. Fonte: Treccani https://www.treccani.it/vocabolario/neo-modello-linguistico-di-grandi-dimensioni_(Neologismi).
ML: Sottoinsieme dell'Intelligenza artificiale che ha come scopo la creazione di algoritmi in grado di apprendere dai dati o migliorare le proprie performance con l'esperienza. Fonte: ISO (trad. Autori, cfr. ed. orig.: https://www.iso.org/artificial-intelligence/machine-learning#toc1).
NLP: Implementazione dell'Intelligenza artificiale nel campo della processazione automatica delle informazioni scritte o parlate in una lingua naturale. Fonte: ISO (trad. aut., cfr. ed. orig.: https://www.iso.org/artificial-intelligence/natural-language-processing).
NER (Named Entity Recogniton) (Li *et al.* 2020).
OpenAI è un laboratorio di ricerca sull'Intelligenza artificiale costituito dalla società no-profit OpenAI, Inc. e dalla sua sussidiaria for-profit OpenAI, L.P. L'obiettivo della ricerca è promuovere e sviluppare un'Intelligenza artificiale user-friendly in modo che l'umanità possa trarne beneficio. ChatGPT è uno dei prodotti della OpenAI.
RAG (Retrieval-Augmented Generation): Tecnica che combina il recupero delle informazioni da un corpus di dati esterno al modello con la generazione di testo per migliorare le risposte dell'IA (Lewis *et al.* 2020).
RLHF (Reinforcement Learning with Human Feedback): L'apprendimento per rinforzo con feedback umano (trad. Autori).
XAI (Explainable AI): campo dell'IA che mira a rendere trasparenti e comprensibili le decisioni dei modelli complessi (Gunning *et al.* 2019).

Bibliografia

Agid (2024), *Strategia italiana per l'Intelligenza Artificiale 2024–2026*, Roma, Agenzia per l'Italia Digitale

Alkaissi H., McFarlane S.I. (2023), Artificial hallucinations in ChatGPT: implications in scientific writing, *Cureus*, 15, n.2, e35179

Ansel J., Yang E., He H., Gimelshein N., Jain A., Voznesensky M., Bao B., Bell P., Berard D., Burovski E., Chauhan G., Chourdia A., Constable W., Desmaison A., DeVito Z., Ellison E., Feng W., Gong J., Gschwind M., Hirsh B., Huang S., Kalambarkar K., Kirsch L., Lazos M., Lezcano M., Liang Y., Liang J., Lu Y., Luk C., Maher B., Pan Y., Puhrsch C., Reso M., Saroufim M., Siraichi M.Y., Suk H., Suo M., Tillet P., Wang E., Wang X., Wen W., Zhang S., Zhao X., Zhou K., Zou R., Mathews A., Chanan G., Wu P., Chintala S. (2024), PyTorch 2: faster machine learning through dynamic Python bytecode transformation and graph compilation, in *ASPLOS '24: Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2*, New York, Association for Computing Machinery, pp.929-947

Augenstein I., Baldwin T., Cha M., Chakraborty T., Ciampaglia G.L., Corney D., Derczynski L., Dutta S., Ghosh S., Gorrell G., Li J., Martino G.D.S., Maynard D., Mittal S., Moens M.F., Nakov P., Pavlopoulos J., Popat K., Rayson P., Schneider J., Thorne J., Zagni G. (2023), Factuality challenges in the era of large language models, *arXiv preprint* <https://doi.org/10.48550/arXiv.2310.05189>

Bolognesi F. (2024), *Intelligenza Artificiale. Istruzioni per l'uso*, Roma, EPC Editore

Boselli R., Cesarini M., Mercorio F., Mezzanzanica M. (2017), Using machine learning for labour market intelligence, in *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2017, Skopje, Macedonia, September 18–22, 2017, Proceedings, Part III*, 10, pp.330-342, Cham, Springer International Publishing

Cazzaniga M., Jaumotte F., Li L., Melina G., Panton A.J., Pizzinelli C., Rockall E.J., Mendes Tavares M. (2024), *Gen-AI: Artificial Intelligence and the Future of Work*, IMF Staff Discussion Note SDN2024/001, Washington, DC, International Monetary Fund

Chen Y., Fang H., Zhao Y., Zhao Z. (2024), *Recovering overlooked information in categorical variables with LLMs: an application to labor market mismatch*, NBER Working Paper n.32327, Cambridge, MA, National Bureau of Economic Research

Excelsior-Unioncamere (2023), *ITS Academy e lavoro. Gli sbocchi lavorativi per la formazione terziaria ITS Academy nelle imprese. Indagine 2023*, Roma, Unioncamere

Feuerriegel S., Hartmann J., Janiesch C., Zschech P. (2024), Generative AI, *Business & Information Systems Engineering*, 66, n.1, pp.111-126

Gunning D., Stefik M., Choi J., Miller T., Stumpf S., Yang G.-Z. (2019), XAI – Explainable artificial intelligence, *Science Robotics*, 4, eaay7120 <https://doi.org/10.1126/scirobotics.aay7120>

Ivanov T., Penchev V. (2024), AI benchmarks and datasets for LLM evaluation, *arXiv preprint* <https://doi.org/10.48550/arXiv.2412.01020>

Lewis P., Perez E., Piktus A., Petroni F., Karpukhin V., Goyal N., Küttler H., Lewis M., Yih W.-T., Rocktäschel T., Riedel S., Kiela D. (2020), Retrieval-augmented generation for knowledge-intensive NLP tasks, *Advances in Neural Information Processing Systems*, 33, pp.9459-9474

Li J., Sun A., Han J., Li C. (2020), A survey on deep learning for named entity recognition, *IEEE Transactions on Knowledge and Data Engineering*, 34, n.1, pp.50-70

Lin J., Tang J., Tang H., Yang S., Chen W.M., Wang W.C., Zhang Y., Han S. (2024), AWQ: activation-aware weight quantization for on-device LLM compression and acceleration, *Proceedings of Machine Learning and Systems*, 6, pp.87-100

Lin S., Hilton J., Evans O. (2021), TruthfulQA: measuring how models mimic human falsehoods, *arXiv preprint* <https://doi.org/10.48550/arXiv.2109.07958>

Singh C. (2023), Machine learning in pattern recognition, *European Journal of Engineering and Technology Research*, 8, n.2, pp.63-68

Sofronic B. (2022), *Sophia. Documentazione tecnica del progetto. Nota tecnica*, Roma, Inapp <https://oa.inapp.org/xmlui/handle/20.500.12916/3619>

Sutton R.S., Barto A.G. (2018), *Reinforcement learning: An introduction*, 2ª ed., Cambridge, MA, MIT Press

Ton J.F., Taufiq M.F., Liu Y. (2024), Understanding chain-of-thought in LLMs through information theory, *arXiv preprint* <https://doi.org/10.48550/arXiv.2411.11984>

Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser Ł., Polosukhin I. (2017), Attention is all you need, in *Advances in Neural Information Processing Systems*, *31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA* <https://arxiv.org/abs/1706.03762v7>

Wang C., Liu X., Yue Y., Tang X., Zhang T., Cheng J., Yao Y., Gao W., Hu X., Qi Z., Wang Y., Yang L., Wang J., Xie X., Zhang Z., Zhang Y. (2023), Survey on factuality in large language models: knowledge, retrieval and domain-specificity, *arXiv preprint* <https://doi.org/10.48550/arXiv.2310.07521>

Alessia Romito
a.romito@inapp.gov.it
Collaboratore di ricerca presso la Struttura Sistemi formativi dell'Inapp. Si occupa di istruzione e formazione professionale e di competenze chiave per l'occupabilità nella filiera lunga della formazione tecnico-professionale. Tra le pubblicazioni recenti si segnala: *L'evoluzione del mercato del lavoro del comparto sanitario nel contesto della digitalizzazione dei servizi e delle prestazioni*, Inapp, WP 103, 2023; *Modelli e prassi di apprendistato formativo a confronto – Studi di caso in Italia ed Europa*, Inapp Report 37, 2023.

Boris Sofronic
b.sofronic@inapp.gov.it
Collaboratore di ricerca presso la Struttura Lavoro e professioni dell'Inapp. Si occupa della Labour Market Intelligence e dell'Intelligenza artificiale, con particolare attenzione al tema del mismatch delle competenze. Tra le pubblicazioni sul tema è coautore di *Verso l'interoperabilità. Una sperimentazione di interconnessione tra strumenti e piattaforme ICT dell'Inapp*, Inapp WP 138, 2025) e autore del Focus *Sfide e opportunità della Labour Market Intelligence* (LMI), Rapporto Inapp 2021.

Mind the gap

L'IA come opportunità di risposta alla domanda di lavoro insoddisfatta in Italia (2024-2033)

Giovanni Armillotta
Randstad Research Institute

Emilio Colombo
Randstad Research Institute

Martina Gnudi
Randstad Research Institute

Francesca Lettieri
Randstad Research Institute

Federica Romano
Randstad Research Institute

Francesco Trentini
Randstad Research Institute

Questo studio analizza l'impatto dell'IA sul mercato del lavoro italiano in chiave prospettica, esplorando la correlazione con i mismatch futuri per singola professione. Vengono costruiti indicatori di *unmet demand* che individuano le professioni con fabbisogno eccedente rispetto all'offerta attesa. Confrontandoli con l'esposizione all'IA, si analizza se le professioni più richieste siano anche le più esposte. Una correlazione positiva suggerisce che l'IA possa sostenere crescita e riduzione dei mismatch.

This study analyzes the impact of AI on the Italian labor market from a forward-looking perspective, exploring its correlation with future mismatches by occupation. We develop unmet demand indicators to identify occupations with expected labor shortages. By comparing these with AI exposure levels, we assess whether the most in-demand occupations are also the most exposed. A positive correlation would suggest that AI can support labor market growth and help reduce skill mismatches.

DOI: 10.53223/Sinappsi_2025-02-15

Citazione
Armillotta G., Colombo E., Gnudi M., Lettieri F., Romano F., Trentini F. (2025), Mind the gap. L'IA come opportunità di risposta alla domanda di lavoro insoddisfatta in Italia (2024-2033), *Sinappsi*, XV, n.2, pp.228-242

Parole chiave
Intelligenza artificiale
Mercato del lavoro
Skill mismatch

Keywords
Artificial intelligence
Labour market
Skill mismatch

Introduzione

La crescente diffusione di tecnologie basate sull'Intelligenza artificiale (IA) sta trasformando profondamente il mondo del lavoro, ridefinendo i processi produttivi, influenzando la domanda di skill e competenze in molte professioni.

In parallelo, si è sviluppato un ampio dibattito pubblico sull'impatto che l'introduzione di questi nuovi strumenti avrà sull'occupazione. Il dibattito si articola attorno a due dinamiche fondamentali: da un lato, la capacità dell'IA di fungere da complemento alle attività umane, potenziando produttività ed efficienza, dall'altro quella di sostituire interamente alcune specifiche funzioni e attività lavorative, riducendo la domanda di specifici profili professionali. Comprendere in quali settori e per quali mansioni l'IA assumerà l'uno o l'altro ruolo è cruciale per anticipare gli effetti sulla struttura occupazionale e adottare strategie adeguate di formazione e politica del lavoro. La letteratura economica ha cercato di misurare e stimare il potenziale impatto dell'IA sul mercato del lavoro. L'approccio maggiormente seguito è stato il quadro analitico del *task-based approach* (Autor 2013) in cui ogni singola professione viene suddivisa in un numero di componenti (task) specifici, per ognuno dei quali viene valutata l'esposizione all'IA costruendo così un indice di esposizione per la

singola professione. Particolarmente significativo è il lavoro di Felten *et al.* (2021) che, basandosi sul repertorio americano di professioni e competenze O*Net, ha assegnato un punteggio di esposizione all'IA attraverso una rilevazione su un campione di lavoratori reclutati tramite Amazon Mechanical Turk. Questa metrica, chiamata AI Occupational Exposure (AIOE), è stata successivamente adattata per cogliere il grado di complementarità e sostituzione delle occupazioni ed è stata applicata in contesti internazionali (Pizzinelli *et al.* 2023) e specificamente a quello italiano (Ferri *et al.* 2024). Un approccio simile è seguito da Eloundou *et al.* (2024) che hanno valutato l'impatto potenziale dei *Large language models* (LLM) e delle tecnologie correlate, proponendo un framework che ne analizza la rilevanza rispetto ai compiti specifici svolti dai lavoratori. La loro ricerca evidenzia la necessità di solide valutazioni di impatto sociale e di misure politiche mirate per affrontare gli effetti degli LLM sul mercato del lavoro. I risultati mostrano come l'impatto potenziale degli LLM è presente in quasi tutti i settori, mostrando un'ampia eterogeneità dovuta principalmente al grado di adozione delle innovazioni e l'integrazione all'interno del processo produttivo. Settori come l'elaborazione dati, l'elaborazione delle informazioni e gli ospedali mostrano un'alta esposizione mentre i lavori cognitivi più di routine sono tra le occupazioni più esposte. Un approccio diverso è seguito da Fenoaltea *et al.* (2024) che hanno introdotto l'indice AI Startup Exposure (AISE). In termini generali l'AISE misura l'esposizione delle professioni alle nuove applicazioni *IA-intensive* introdotte da startup. Questo approccio si distingue dai precedenti per il fatto che non ricade nel quadro analitico task-based, bensì mette in relazione le short job description di ciascuna professione in O*Net con i prodotti contenenti tecnologie di IA sviluppati da startup finanziate dall'acceleratore Y Combinator. Ricorrendo a un modello linguistico viene inferita la relazione tra i prodotti e i servizi offerti e la singola professione. I risultati mostrano che, sebbene le professioni altamente qualificate siano teoricamente molto esposte all'IA, le innovazioni introdotte dalle startup le riguardano in modo differenziato. Infatti, il rischio effettivo di sostituzione dipende non solo dalla fattibilità tecnica, ma anche da considerazioni etiche e sociali.

Un confronto tra l'indice di esposizione occupazionale all'IA (AIOE) di Felten *et al.* (2021) e l'Indice AISE è presentato da un recente report della Commissione europea (Albora *et al.* 2025). I risultati relativi all'Indice AIOE, basati sui task che compongono i lavori, mostrano che l'esposizione all'IA è alta per impiegati e lavoratori altamente qualificati. Gli autori mostrano che mettendo in relazione l'Indice AIOE con una misura di similarità tra professioni in termini di competenze che le caratterizzano, si osserva una potenziale 'trappola dell'IA' con scarse opportunità di riqualificazione e ricollocazione per i lavoratori spiazzati dall'introduzione dell'IA. Tuttavia, analizzando lo sviluppo effettivo dell'IA da parte delle startup tramite l'Indice AISE, si osserva che l'impatto dell'IA è eterogeneo tra i lavori cognitivi, indicando che non tutti i lavori sono ugualmente sostituibili nonostante le abilità comuni. Parallelamente, Simons *et al.* (2024) discutono le implicazioni economiche e politiche dei recenti progressi dell'Intelligenza artificiale, distinguendo tra i modelli di IA predittiva, già in uso da tempo, e i più recenti modelli di IA generativa, capaci di creare contenuti originali. Nonostante il grande interesse mediatico, il documento sottolinea come la diffusione e l'adozione effettiva dell'Intelligenza artificiale nell'UE rimangano ancora limitate a pochi settori e, prevalentemente, a grandi aziende. Mentre la maggior parte dei lavori citati si occupa di fornire una metodologia per misurare l'impatto dell'IA e il grado di esposizione della forza lavoro attuale, la relazione tra l'esposizione dei diversi profili professionali all'IA e il fabbisogno occupazionale *futuro* risulta un ambito ancora relativamente poco esplorato. Se l'IA è impiegata come complemento, i lavoratori possono beneficiare di strumenti avanzati che ne amplificano le capacità con un conseguente aumento della produttività e un corrispondente aumento della domanda di competenze specifiche per il suo utilizzo. Al contrario nei casi in cui l'IA opera come sostituto, si può generare una contrazione dell'occupazione in determinati segmenti del mercato del lavoro. Mentre l'aumento della produttività è sempre valutato positivamente, è in relazione ai fabbisogni occupazionali non soddisfatti che può avere un effetto positivo anche l'introduzione di soluzioni tecnologiche che sostituiscano posti di lavoro, risultando in un impatto positivo sul benessere collettivo. Occorre considerare, infatti, che l'innovazione tecnologica e l'IA si

inseriscono in un contesto peculiare, caratterizzato dall'invecchiamento della popolazione, dovuto alla combinazione di bassa natalità e aumento dell'aspettativa di vita, che sta determinando una contrazione della forza lavoro e un aumento dei tassi di dipendenza. La pressione sul mercato del lavoro che deriva da queste dinamiche è duplice in quanto la riduzione nel numero di individui attivi comporta parallelamente un aumento di individui da destinare all'assistenza. Sia che questa funzione sia svolta da professionisti sia che sia svolta dalle reti familiari, l'effetto è una riduzione della forza lavoro disponibile, in termini di numero di occupati o di ore lavorate. In questo contesto, dunque, anche se le nuove tecnologie dovessero sostituire forza lavoro dove essa viene a mancare per ragioni demografiche, contribuirebbero a risolvere un problema anziché crearne uno. L'obiettivo di questo lavoro è di contribuire alla letteratura relativa all'impatto del cambiamento tecnologico e in particolare dell'IA sull'occupazione, collegando esposizione all'IA e fabbisogno occupazionale insoddisfatto in Italia negli anni 2024-2033. Il lavoro si apre con la presentazione delle caratteristiche del fabbisogno occupazionale in Italia, declinato nell'aspetto della domanda di rimpiazzo (*replacement demand*) e di espansione (*expansion demand*) e a livello di gruppi occupazionali, in ottica comparata europea. In particolare, si espongono le evidenze relative alle dinamiche dell'Unione europea e di quattro economie chiave: Italia, Francia, Spagna e Germania. Questo passaggio preliminare consente di comprendere se e in che misura l'evoluzione del mercato del lavoro italiano segua tendenze analoghe ad altri Paesi europei o se, al contrario, si evidenzino specificità nazionali che potrebbero influenzare diversamente il rapporto tra IA e occupazione. Nella seconda fase dell'indagine presentiamo un'analisi dettagliata del caso italiano, presentando evidenze circa la relazione tra esposizione all'IA e domanda insoddisfatta per il periodo 2024-2033 con un livello di dettaglio che corrisponde alla classificazione Istat dei codici professionali a 4 digit. Tale metodo consente di valutare in modo sistematico se le occupazioni per cui è prevista una maggiore richiesta di lavoratori siano anche quelle più esposte all'IA. Una correlazione positiva tra il fabbisogno occupazionale e l'esposizione all'IA implica che quest'ultima può rappresentare un elemento di supporto alla crescita del mercato del lavoro italiano, contribuendo a ridurre il mismatch tra competenze richieste e competenze disponibili.

I risultati contribuiscono a fornire una chiave di lettura innovativa delle conseguenze sull'occupazione dell'introduzione delle tecnologie di IA nei contesti lavorativi. Questo quadro analitico e l'interpretazione delle evidenze empiriche che determina offrono spunti per un nuovo orientamento delle politiche di sviluppo delle risorse umane, un'integrazione efficace della tecnologia nei contesti lavorativi.

1. Metodologia e dati

L'approccio metodologico prevede una serie di passaggi volti a garantire la comparabilità tra i diversi dataset e un'analisi accurata delle tendenze occupazionali nel periodo 2024-2033. Per analizzare il fabbisogno occupazionale sia in Italia che, in un'ottica comparativa, nei principali Paesi europei sono utilizzate le previsioni effettuate dall'Agenzia europea Cedefop che ha sviluppato un modello di lungo periodo per prevedere il fabbisogno occupazionale per ciascun Paese europeo sino al 2035[1]. Seguendo un'impostazione consolidata, le previsioni del fabbisogno elaborate dal Cedefop sono frutto del combinato disposto di due elementi: l'*expansion* e la *replacement demand*. L'*expansion demand*, o domanda di espansione, misura la crescita netta dell'occupazione derivante dall'aumento della domanda di lavoro in determinati settori o professioni, e viene quindi misurata come variazione annuale dello stock di occupati. L'*expansion demand* può essere per definizione sia di segno positivo che negativo, a seconda che l'occupazione settoriale cresca o diminuisca. Le stime Cedefop sono effettuate sotto l'ipotesi che le dinamiche occupazionali seguano caratteristiche spiccatamente settoriali. L'*expansion demand* costituisce tuttavia solo una parte del fabbisogno complessivo: anche in settori in crisi, nei quali si verifica una contrazione complessiva dei livelli di impiego, vi sono infatti opportunità di lavoro che si aprono. In altri termini occorre considerare un'ulteriore componente della domanda di lavoro, la cosiddetta *replacement demand* o domanda di rimpiazzo. Questa è costituita dalla domanda di lavoratori necessari a compensare

1 Le previsioni sono disponibili presso: https://www.cedefop.europa.eu/en/tools/skills-forecast.

i posti lasciati vacanti a causa di pensionamenti, mortalità o mobilità intersettoriale. A differenza dell'*expansion demand*, la *replacement demand* è sempre positiva e, poiché si riferisce all'intero stock della popolazione lavorativa, risulta ampiamente superiore all'altra componente. Il fabbisogno occupazionale totale risulta dunque dalla somma algebrica di queste due componenti. Dal momento che le previsioni Cedefop sono espresse al secondo livello ISCO-08, sono stati utilizzati i microdati della Forze di Lavoro Istat per stimare la distribuzione del fabbisogno al quarto digit CP2011. La procedura è differenziata per *replacement* ed *expansion demand*. Nel primo caso, la composizione della domanda è stata inferita dalla composizione degli occupati che nel 2023 si trovano nella fascia tra i 55 e i 64 anni, ossia quelli destinati a uscire dal mercato del lavoro negli anni di interesse. Più in dettaglio, sono stati estratti dalla banca dati Istat delle Forze di Lavoro il numero di occupati nel 2023 a livello CP2011 4 digit per questa fascia di età. Per allineare questi dati con le stime previsionali del Cedefop, è stata prima ricavata la corrispondenza tra CP2011 4 digit e ISCO-08 2 digit e, successivamente, calcolata la distribuzione degli occupati CP2011 4 digit all'interno di ciascun ISCO-08 2 digit, attribuendo così la quota della *replacement demand* complessiva tra il 2024 e il 2033 per ogni CP2011 4 digit. La stima segue un criterio diverso per la seconda componente, l'*expansion demand*, che, ricordiamo, misura la variazione netta dell'occupazione dovuta alla crescita economica e settoriale. In questo caso si è utilizzata come riferimento la distribuzione degli occupati di età 15-54 anni nel 2023. L'ipotesi implicita, e principale limitazione di questo schema analitico, che si applica a entrambe le componenti di domanda è che la distribuzione occupazionale al 4 livello rifletta in forma costante la quota di occupazione per l'intero periodo considerato. Tuttavia, si tratta anche dell'ipotesi più conservativa, che permette di rappresentare uno scenario di riferimento a tecnologia e mercato del lavoro invariati, facilitando l'interpretazione dei risultati. In questo modo, abbiamo ottenuto una stima dettagliata del fabbisogno totale per ogni profilo professionale a livello CP2011 4 digit, consentendo un'analisi più fine dell'impatto dell'IA sulle diverse categorie occupazionali.

L'analisi sopra descritta si riferisce al lato della domanda di lavoro da parte del settore produttivo. Per poter quindi caratterizzare la domanda insoddisfatta è necessario stimare anche la componente dell'offerta. In questo caso si è calcolata la componente dell'offerta di lavoro, considerando come riferimento la popolazione tra i 25 e i 34 anni. Sotto l'ipotesi che nel periodo di riferimento 2024-2033 la struttura occupazionale di questo gruppo di occupati rimanga invariata, abbiamo applicato questa distribuzione alle previsioni dei nuovi occupati nel medesimo periodo. In particolare, abbiamo considerato la previsione della popolazione tra i 25 e i 34 anni per il periodo 2024-2033, calcolato la media, e moltiplicato i tassi di occupazione attuali di tutti i CP2011 4 digit per questo valore, ottenendo così la componente dell'offerta nel periodo di interesse per ciascun CP2011 4 digit.

Sia considerando il lato della domanda sia quello dell'offerta, la nostra metodologia presenta elementi di criticità. In particolare, assumiamo che la distribuzione delle professioni al 4 livello all'interno delle professioni al 2 livello rimanga invariata così come rimanga invariata la distribuzione delle occupazioni dei giovani in entrata nel mondo del lavoro. Entrambe le ipotesi sono frutto della scarsa disponibilità di dati sufficientemente granulari per poter effettuare un'analisi maggiormente robusta, ma sono ovviamente discutibili. Le scelte fatte offrono tuttavia un benchmark conservativo su cui effettuare le analisi. In altri termini analizziamo domanda e offerta di lavoro ipotizzando costante la struttura occupazionale attuale. Per concludere, abbiamo calcolato il fabbisogno insoddisfatto, o *unmet demand*, per il periodo 2024-2033 a livello di singola professione CP2011 4 digit come saldo tra domanda totale, comprendente *replacement* ed *expansion demand*, e offerta di lavoro.

Questo passaggio è fondamentale per comprendere quali professioni saranno più richieste nel prossimo futuro e per identificare le professioni per cui la richiesta di lavoratori potrebbe superare o, al contrario, essere inferiore alla disponibilità di nuove risorse nel mercato del lavoro. Con riferimento all'esposizione all'IA, in questo lavoro viene utilizzato l'indicatore elaborato da Felten *et al.* (2021) che assegna un punteggio di esposizione all'IA a ciascun elemento della Standard Occupational Classification (SOC) statunitense sulla base del repertorio O*Net. Questo indicatore misura il grado di esposizione di una professione all'Intelligenza artificiale, con particolare riferimento a mansioni non ripetitive

e cognitive, includendo anche l'interazione con sistemi robotici, ma non limitandosi ad essa. Gli autori quantificano questa esposizione attraverso la costruzione dell'AI Occupational Exposure (AIOE), basato sull'associazione tra 10 applicazioni comuni dell'IA e 52 abilità lavorative definite nel database americano O*Net. Per ogni abilità, viene calcolato un punteggio di esposizione all'IA sommando i valori di relazione tra l'abilità stessa e ciascuna delle 10 applicazioni di IA considerate. L'Indice AIOE costituisce uno standard di riferimento in letteratura. La sua ampia diffusione, unita all'attualità dei dati, garantisce infatti un'elevata comparabilità con altre ricerche sul tema. Poiché questa procedura è stata originariamente sviluppata per il mercato del lavoro statunitense, l'adattamento al contesto italiano ha richiesto una serie di operazioni di transcodifica: prima dal codice SOC (Standard Occupational Classification) al codice ISCO-08 a 4 digit, e successivamente dal codice ISCO-08 a 4 digit ai codici professionali CP2011 a 4 digit utilizzati dall'Istat. L'ultimo passaggio della nostra analisi consiste nell'esaminare la relazione tra la domanda insoddisfatta, per il periodo 2024-2033, e il grado di esposizione all'Intelligenza artificiale. In questo modo è possibile evidenziare quali professioni vedano associati alti livelli di esposizione all'IA e alta domanda insoddisfatta, per le quali la sostituibilità con l'IA potrebbe essere una leva positiva di intervento proprio per colmare la domanda insoddisfatta.

2. Risultati

La tabella 1 analizza la domanda di lavoro nei principali Paesi europei e in UE nel periodo 2024-2033, soffermandosi su tre indicatori chiave: il tasso di *replacement*, il tasso di *expansion* e il tasso di fabbisogno.

Il tasso di *replacement*, che misura la necessità di sostituire i lavoratori in uscita, è particolarmente elevato in Spagna, con un valore del 36,6%, seguita dalla Germania al 32,8% e dall'Italia al 31,8%. Questi dati suggeriscono che in questi Paesi una quota consistente della forza lavoro dovrà essere rimpiazzata nei prossimi anni. La Francia, con un valore del 29,0%, mostra un turnover più contenuto rispetto alla media UE, che si attesta al 31,3%.

Per quanto riguarda il tasso di *expansion*, che stima la crescita netta dell'occupazione, si evidenzia un forte incremento in Spagna, (+11,5%). Questo dato, apparentemente un *outlier*, va interpretato alla luce di dinamiche specifiche del contesto spagnolo. Il fattore predominante è l'impatto della riforma del mercato del lavoro del 2021, che ha causato una massiccia conversione di contratti da temporanei a stabili. Questo fenomeno, pur rappresentando una cruciale stabilizzazione di lavoro già esistente, viene registrato statisticamente come una forte espansione di posti di lavoro permanenti. A questo shock strutturale si sommano un robusto rimbalzo economico post-pandemico, specialmente nel settore dei servizi, e l'impulso degli investimenti legati ai fondi Next Generation EU.

L'UE presenta un tasso moderato del 3,3%, mentre Italia e Francia si attestano su valori simili, rispettivamente al 3,0% e al 2,8%.

La Germania, invece, presenta un quadro unico, registrando un tasso di *expansion* negativo del -0,5%. Questa lieve contrazione netta dell'occupazione non è solo un segnale di rallentamento economico, ma riflette sfide più profonde. Da un lato, l'economia tedesca risente della crisi del suo settore manifatturiero, penalizzato dagli alti costi energetici e dalla debole domanda globale. Dall'altro, il Paese affronta una profonda transizione strutturale (come nell'automotive) e, soprattutto, un rilevante vincolo demografico. La massiccia ondata di pensionamenti, unita alla carenza di manodopera qualificata, fa sì che molte posizioni lavorative non vengano sostituite, contribuendo a una distruzione di posti di lavoro per attrito che supera la capacità del sistema di crearne di nuovi.

In conclusione, al netto delle variazioni occupazionali dovute alla riforma del 2021, la Spagna si conferma il Paese con la più alta domanda di lavoratori nel contesto europeo, una dinamica che poggia su due pilastri fondamentali: una robusta crescita economica e un mercato del lavoro eccezionalmente dinamico.

Da un lato, l'economia spagnola sta sovraperformando i principali partner europei. Dopo una ripresa post-pandemica consolidata, il Paese continua a mostrare tassi di crescita del PIL superiori alla media UE, crescita trainata da un settore dei servizi vitale e da un turismo che ha raggiunto livelli record. Questa forte espansione economica alimenta una domanda diretta di nuovi posti di lavoro (la cosiddetta domanda per 'crescita netta').

Dall'altro lato, a questa domanda si affianca un'offerta di lavoro in espansione che vede entrare

Tabella 1. Tasso di replacement demand, expansion demand e fabbisogno dell'UE e di alcuni Paesi europei nell'arco temporale 2024-2033

Paesi	Tasso di replacement	Tasso di expansion	Tasso di fabbisogno
UE 27	31,3%	3,3%	34,6%
Italia	31,8%	3,0%	34,8%
Francia	29,0%	2,8%	31,8%
Germania	32,8%	-0,5%	32,3%
Spagna	36,6%	11,5%	48,1%

Fonte: elaborazioni Randstad Research su dati Cedefop

nel mercato del lavoro una percentuale sempre maggiore della popolazione in età lavorativa, perché incoraggiata da un clima di fiducia e da maggiori opportunità. In un'Europa che invecchia, la Spagna si distingue proprio per l'aumento del tasso di partecipazione, soprattutto femminile (Istat 2024). Questo significa che, a differenza di altri Paesi con demografia più stagnante, l'economia spagnola può attingere a un bacino di manodopera aggiuntivo per sostenere la crescita.

Infine, a questa domanda generata dalla crescita economica e alimentata da una maggiore partecipazione, si aggiunge l'elevato fabbisogno di rimpiazzo (domanda per 'sostituzione') dei lavoratori della generazione dei baby boomer in uscita per pensionamento. La combinazione di questi fattori spiega perché la Spagna presenti la domanda di lavoro più intensa d'Europa.

La Germania, invece, mostra segnali di contrazione, con una riduzione del numero di occupati e un tasso di *expansion* negativo. Italia e Francia presentano dinamiche simili, caratterizzate da un elevato tasso di *replacement* ma da una crescita più contenuta. Nel complesso, nei prossimi dieci anni in UE si prevede un fabbisogno pari a più di un terzo della forza lavoro. Questa cifra dà l'idea del grande cambiamento che attraverserà il mercato del lavoro europeo sia in termini numerici sia in termini di competenze necessarie per affrontare le sfide tecnologiche poste dall'IA.

La tabella 2 presenta un'analisi comparativa dei tassi di *replacement* e di *expansion* per Italia, Francia, Germania e Spagna, con i valori espressi come differenza rispetto alla media dell'Unione europea. L'analisi è condotta a livello dei profili professionali ISCO-08 a 2 digit.

L'Italia, soprattutto per i primi due grandi gruppi professionali, mostra differenze positive significative, segnalando un tasso di *replacement* più elevato rispetto alla media UE e suggerendo una maggiore necessità di sostituire i lavoratori in uscita per queste specifiche categorie professionali. Ad esempio, si osserva che per i profili ISCO-08 12 (Administrative and commercial managers) e ISCO-08 22 (Health professionals) l'Italia mostra una differenza positiva marcata, indicando un fabbisogno di sostituzione superiore alla media UE per questa categoria. Tuttavia, per altri profili, le differenze sono minime o addirittura negative, segnalando un tasso di *replacement* in linea o inferiore alla media UE. Per quanto riguarda i tassi di *expansion*, le differenze variano a seconda dei profili ISCO, con alcuni profili che mostrano differenze positive, indicando una crescita dell'occupazione superiore alla media UE, mentre altri mostrano differenze negative, suggerendo una crescita inferiore o addirittura una contrazione rispetto alla media. Ad esempio, per il profilo ISCO 14 (Hospitality, retail and other services managers) l'Italia mostra una differenza positiva elevata nel tasso di *expansion*, mentre per il profilo ISCO 42 (Customer services clerks), la differenza risulta negativa.

La Francia, in generale, tende a mostrare differenze negative nel tasso di *replacement* rispetto alla media UE, suggerendo che il fabbisogno di sostituzione dei lavoratori in uscita è inferiore rispetto alla media europea, il che potrebbe riflettere una maggiore stabilità del mercato del lavoro francese in alcuni settori o una diversa struttura demografica della forza lavoro. In alcuni casi, però, si notano delle eccezioni: per il profilo ISCO-08 11 (Chief executives, senior officials and legislators) la Francia presenta una differenza positiva elevata nel tasso di *replacement*. Le differenze nei tassi di *expansion* sono eterogenee, con alcuni profili che mostrano una crescita dell'occupazione superiore alla media UE, mentre altri presentano valori inferiori,

Tabella 2. Differenze tassi di replacement e expansion tra i Paesi europei e l'Unione europea 2023-2033

Isco 2 digit	Descrizione ISCO 2 digit	Italia		Francia		Germania		Spagna	
		Tasso di replacement	Tasso di expansion	Tasso di replacement	Tasso di expansion	Tasso di replacement	Tasso di expansion	Tasso di replacement	Tasso di expansion
11	Chief executives, senior officials and legislators	0,01	-0,08	0,16	-0,07	0,01	-0,05	-0,03	-0,11
12	Administrative and commercial managers	0,21	0,04	-0,07	-0,22	0,04	0,04	0,04	0,12
13	Production and specialised services managers	0,12	-0,10	-0,01	0,33	0,00	-0,12	0,07	-0,08
14	Hospitality, retail and other services managers	0,14	0,31	-0,09	-0,10	0,00	-0,21	0,01	-0,13
21	Science and engineering professionals	0,03	0,09	-0,01	0,06	0,03	-0,12	-0,02	0,03
22	Health professionals	0,23	-0,03	0,00	-0,08	-0,01	-0,09	0,05	-0,11
23	Teaching professionals	0,12	0,09	-0,06	-0,08	0,00	0,06	0,05	0,10
24	Business and administration professionals	0,12	0,17	-0,01	0,00	-0,02	-0,10	0,09	0,04
25	Information and communications technology professionals	-0,03	-0,05	0,02	0,04	-0,04	-0,18	0,06	0,08
26	Legal, social and cultural professionals	-0,01	0,03	0,00	-0,04	0,04	-0,03	0,02	-0,06
31	Science and engineering associate professionals	-0,03	0,04	-0,08	-0,08	0,04	-0,05	0,02	0,14
32	Health associate professionals	0,00	0,02	-0,02	0,06	0,00	-0,04	0,06	0,16
33	Business and administration associate professionals	0,01	-0,01	-0,02	-0,02	0,01	-0,03	0,09	0,05
34	Legal, social, cultural and related associate professionals	-0,02	-0,12	-0,05	-0,07	-0,01	0,06	0,07	0,23
35	Information and communications technicians	-0,03	-0,07	-0,01	-0,07	-0,09	-0,33	0,14	0,22
41	General and keyboard clerks	0,00	0,07	-0,06	0,05	0,07	-0,04		
42	Customer services clerks	-0,05	-0,18	0,02	0,15	-0,01	0,05	0,07	0,13
43	Numerical and material recording clerks	-0,03	-0,02	-0,08	-0,37	-0,02	-0,04	0,07	0,23
44	Other clerical support workers	0,09	0,08	-0,23	-0,19	0,01	0,06	0,16	-0,08
51	Personal service workers	-0,08	-0,04	-0,06	-0,08	0,02	-0,05	0,10	0,21
52	Sales workers	-0,07	-0,06	0,03	-0,02	0,02	-0,05	0,07	0,15
53	Personal care workers	0,08	0,02	-0,06	-0,10	0,01	0,02	0,02	-0,04
54	Protective services workers	-0,07	-0,09	-0,04	0,07	-0,05	0,03	0,03	0,11
61	Market-oriented skilled agricultural workers	0,06	0,17	0,10	0,28	0,13	0,29	0,15	0,27
62	Market-oriented skilled forestry, fishery and hunting workers	0,00	-0,19	-0,05	0,21	0,08	0,37	-0,24	-0,27
71	Building and related trades workers, excluding electricians	0,02	0,06	-0,01	-0,04	-0,01	-0,07	0,05	0,06
72	Metal, machinery and related trades workers	-0,06	0,04	0,01	0,00	-0,01	-0,04	-0,01	0,05
73	Handicraft and printing workers	-0,07	-0,19	0,01	-0,05	0,01	0,07	0,10	-0,01
74	Electrical and electronic trades workers	-0,07	0,05	-0,01	0,08	0,00	-0,07	0,05	0,08
75	Food processing, wood working, garment and other craft and re	0,00	-0,10	0,02	0,27	0,03	-0,02	0,07	0,02
81	Stationary plant and machine operators	-0,05	-0,06	-0,04	0,02	0,03	-0,06	0,04	0,15
82	Assemblers	-0,07	-0,04	-0,04	-0,17	0,09	-0,03	0,04	0,35
83	Drivers and mobile plant operators	-0,05	-0,10	-0,04	0,05	0,11	-0,08	-0,02	0,03
91	Cleaners and helpers	0,03	0,08	-0,01	-0,05	0,09	-0,01	-0,02	-0,05
92	Agricultural, forestry and fishery labourers	0,02	0,05	-0,01	-0,15	0,20	0,14	-0,03	-0,01
93	Labourers in mining, construction, manufacturing and transport	-0,04	0,02	-0,03	-0,20	0,02	-0,05	0,07	0,23
94	Food preparation assistants	-0,09	-0,08	0,05	0,09	0,01	-0,14	0,07	0,28
95	Street and related sales and service workers	-0,04	0,05					-0,38	-0,56
96	Refuse workers and other elementary workers	-0,03	-0,02	0,02	0,23	0,11	-0,07	0,01	0,00

Fonte: elaborazioni Randstad Research su dati Cedefop

riflettendo dinamiche settoriali specifiche. In alcuni casi si evidenziano delle differenze negative importanti, come per il profilo ISCO-08 43 (Numerical and material recording clerks).

La Germania presenta un quadro misto, con alcuni profili che mostrano differenze positive, soprattutto nei grandi gruppi professionali ISCO-08 6 (Skilled agricultural, forestry and fishery workers), ISCO-08 8 (Plant and machine operators and assemblers) e ISCO-08 9 (Elementary occupations), indicando un maggiore bisogno di sostituzione rispetto alla media UE, mentre altri mostrano differenze negative o prossime allo zero, il che potrebbe riflettere le sfide demografiche che la Germania sta affrontando, con un invecchiamento della popolazione attiva in alcuni settori che richiede un maggiore ricambio. In generale, il Paese tende a mostrare differenze negative nei tassi di *expansion* rispetto alla media UE, soprattutto nei primi tre gruppi, coerentemente con le tendenze generali di contrazione del mercato del lavoro tedesco evidenziate dalla tabella 1, il che potrebbe riflettere sfide economiche, cambiamenti strutturali o l'impatto di fattori demografici sull'occupazione.

La Spagna, diversamente dagli altri Paesi, mostra perlopiù differenze positive nel tasso di *replacement* rispetto alla media UE. Questo potrebbe suggerire che, per alcuni profili, il mercato del lavoro spagnolo sperimenta un maggior fabbisogno insoddisfatto o che la necessità di sostituzione è più accentuata rispetto alla media europea. La Spagna, in alcuni casi, mostra differenze positive significative nei tassi di *expansion*, indicando una crescita dell'occupazione più forte rispetto alla media UE, e suggerendo una dinamica del mercato del lavoro più espansiva per alcuni profili

professionali. Una professione per cui la differenza è negativa è il profilo ISCO-08 95 (Street and related sales and service workers).

In generale, l'analisi a livello di profili ISCO-08 a 2 cifre consente di cogliere le specificità settoriali e le diverse dinamiche del mercato del lavoro all'interno di ciascun Paese. Le differenze rispetto alla media UE evidenziano le aree in cui ciascun Paese si discosta dalle tendenze europee, sia in termini di fabbisogno di sostituzione dei lavoratori (*replacement*) sia in termini di crescita dell'occupazione (*expansion*). Queste differenze possono essere influenzate da una varietà di fattori, tra cui la struttura economica, le politiche del lavoro, le tendenze demografiche, i cambiamenti tecnologici e le condizioni macroeconomiche.

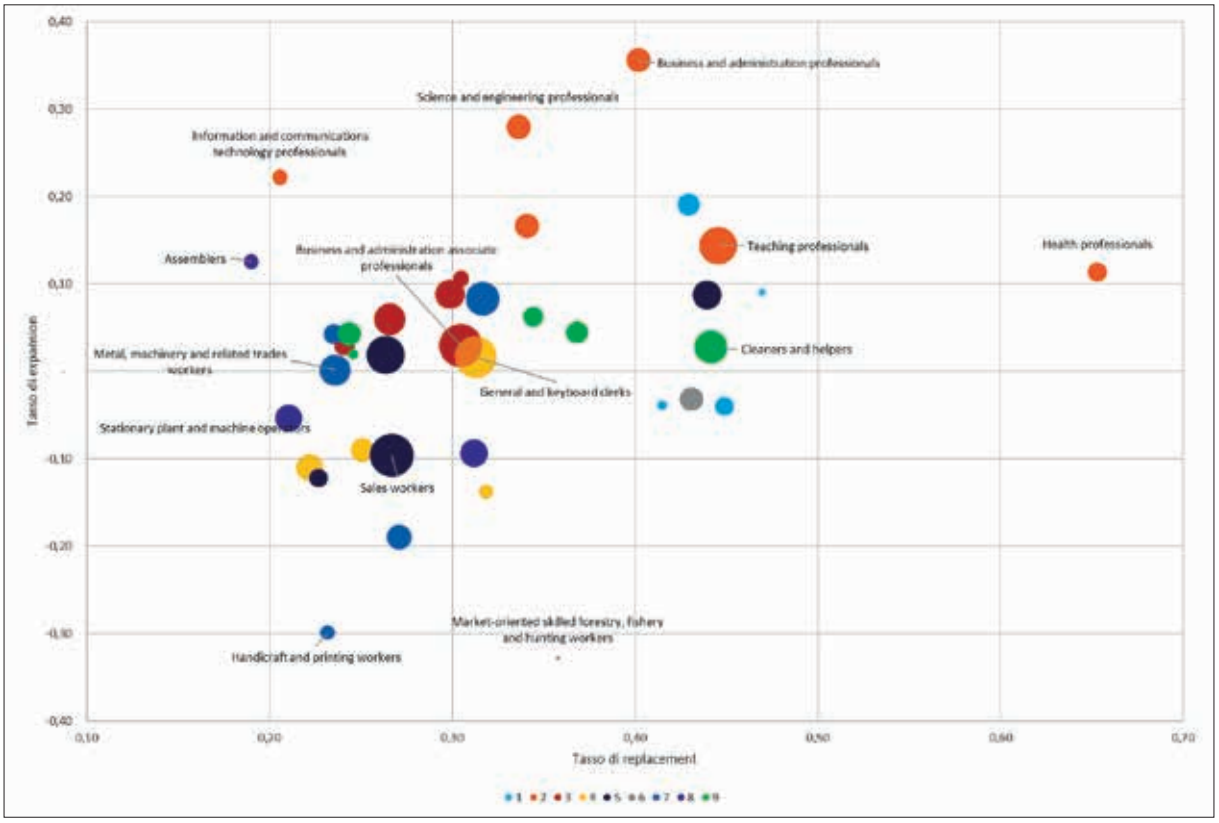
Focalizzando l'attenzione sull'Italia, si mostra la relazione tra il tasso di *replacement* e il tasso di *expansion* per i grandi gruppi professionali ISCO-08 2 digit, evidenziando le differenze nelle dinamiche occupazionali tra i vari settori (figura 1). Le professioni manageriali (ISCO-08 1) presentano un tasso di *replacement* contenuto e un'espansione limitata, riflettendo un mercato del lavoro relativamente stabile per questi ruoli, in cui il turnover è ridotto e la creazione di nuove posizioni non è particolarmente elevata. Le professioni intellettuali, scientifiche e ad alta specializzazione (ISCO-08 2) mostrano invece un'elevata espansione e un tasso di *replacement* variabile, suggerendo un settore in crescita dove il fabbisogno di nuove competenze è significativo e il ricambio generazionale gioca un ruolo chiave. Le professioni tecniche (ISCO-08 3) si collocano in una posizione intermedia, con tassi di *replacement* ed *expansion* moderati, indicando un mercato del lavoro dinamico ma senza le stesse spinte espansive delle professioni altamente qualificate. Gli impiegati (ISCO-08 4) evidenziano un tasso di *expansion* piuttosto basso, a volte negativo, e un turnover moderato, segnalando un settore più statico con scarse nuove opportunità lavorative. Nel settore dei lavoratori dei servizi e del commercio (ISCO-08 5), si osservano tassi di *replacement* variabili e tassi di *expansion* bassi o negativi, riflettendo la continua domanda di lavoratori in questi ambiti, sebbene con alcune differenze tra specifiche professioni. Le professioni agricole e della pesca (ISCO-08 6) presentano generalmente un basso tasso di *expansion*, suggerendo una crescita limitata del settore, mentre il *replacement* indica una sostituzione della forza lavoro che può essere legata a fattori demografici o contrattuali. Gli artigiani e operai specializzati dell'industria ed edilizia (ISCO-08 7) mostrano un tasso di *replacement* moderato con un'espansione bassa o negativa. I conduttori di impianti e macchinari (ISCO-08 8) presentano un comportamento simile, con un turnover visibile ma una crescita occupazionale limitata. Infine, le professioni non qualificate (ISCO-08 9) si distinguono per un tasso di *replacement* variabile e un tasso di espansione leggermente superiore allo zero. Complessivamente, le evidenze qui esposte mostrano che, nel nostro Paese, le professioni ad alta qualificazione (grandi gruppi 1, 2 e 3) tendano ad essere caratterizzati da un tasso di *expansion* e di *replacement* maggiore della media UE. Questi sono dunque i gruppi professionali in cui è atteso il maggior cambiamento nella forza lavoro. Come sottolineato precedentemente le professioni maggiormente qualificate sono anche quelle maggiormente esposte all'IA. In che misura questi due elementi (cambiamento atteso ed esposizione all'IA) avranno un impatto trasformativo positivo nel mercato del lavoro italiano dipenderà anche dalla dinamica dell'offerta di lavoro e da dove si concentreranno i *mismatches*.

La figura 2 mostra la correlazione tra domanda insoddisfatta, o mismatch[2], e l'impatto dell'IA. Per facilitare la lettura del grafico ed evidenziare le tendenze principali a livello aggregato, sono state raggruppate con lo stesso colore le professioni appartenenti allo stesso grande gruppo professionale (da 1 a 9) dei profili CP2011 a 4 digit di Istat.

Il gruppo 1 (Legislatori, imprenditori e alta dirigenza) mostra un impatto dell'IA generalmente positivo e una variabilità relativamente contenuta nel mismatch. Questo suggerisce che, pur essendo soggette a trasformazioni tecnologiche, queste professioni hanno un buon grado di adattabilità, grazie a competenze strategiche e gestionali difficilmente automatizzabili. Le professioni intellettuali, scientifiche e di elevata specializzazione (gruppo 2) evidenziano una presenza consistente nella parte alta del grafico, indicando un impatto significativo dell'IA. Tuttavia, il mismatch varia a

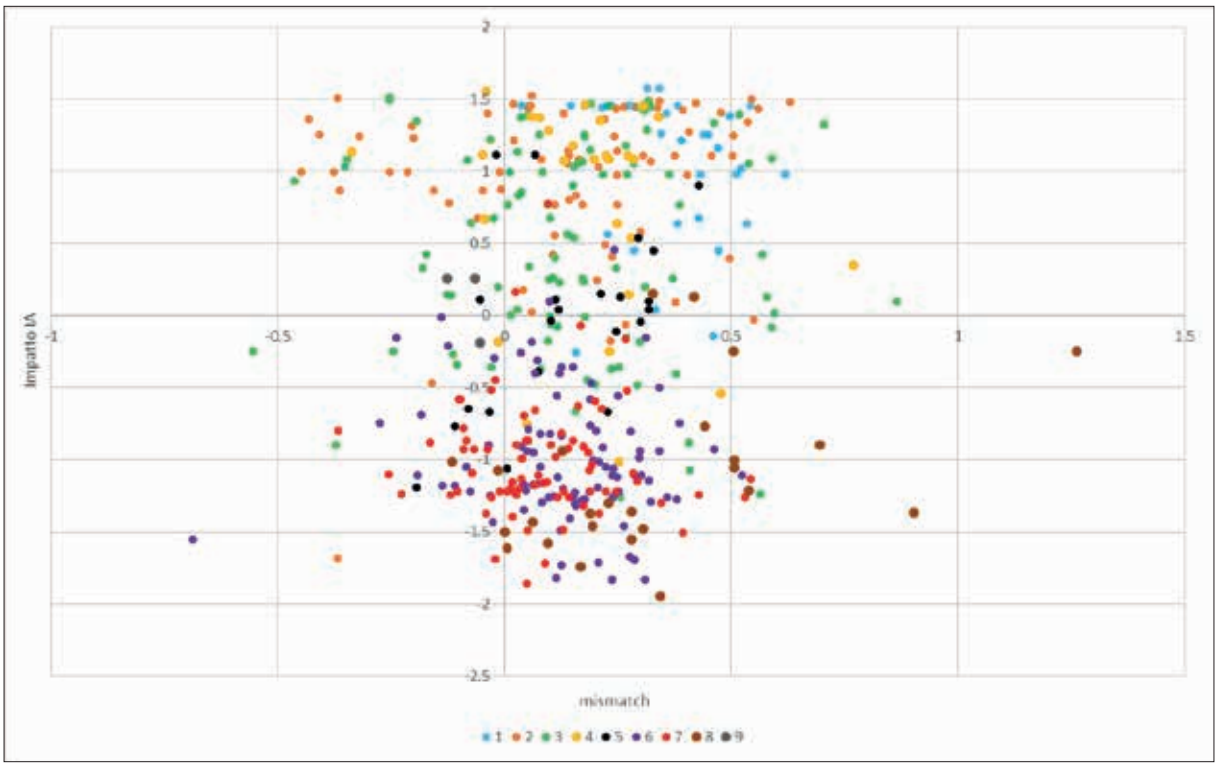
2 Considerata la distribuzione molto asimmetrica della misura di domanda insoddisfatta, al mismatch è stata applicata la seguente trasformazione: lmismatch=sgn(mismatch)*ln(1+|mismatch|).

Figura 1. Tassi di replacement ed expansion per profili professionali ISCO-08 2 digit per l'Italia 2023-2033[3]



Fonte: elaborazioni Randstad Research su dati Cedefop

Figura 2. Relazione tra mismatch ed esposizione all'Intelligenza artificiale sui profili professionali in Italia 2023-2033



Fonte: elaborazioni Randstad Research su dati Cedefop, Rilevazione sulle Forze di lavoro (Istat) e Felten *et al.* (2021)

seconda dei profili. Le professioni tecniche (gruppo 3) si distribuiscono in modo più frammentato, con alcune professioni esposte a un forte impatto dell'IA e un fabbisogno insoddisfatto contenuto, mentre altre evidenziano un disallineamento più marcato. Questa eterogeneità suggerisce che la capacità di adattamento all'IA delle professioni tecniche non è intrinseca alla professione stessa, ma è fortemente influenzata dalle risorse e dalla cultura del settore in cui operano. Ad esempio, settori ad alta tecnologia (come l'ICT o l'automazione industriale) investono costantemente in formazione e integrano rapidamente le nuove competenze, mantenendo basso il mismatch. Al contrario, in settori più tradizionali, la minor propensione all'investimento in riqualificazione può generare un divario di competenze più ampio e, di conseguenza, un maggior disallineamento.

Le professioni del gruppo 4 (Professioni esecutive nel lavoro d'ufficio) tendono a concentrarsi in aree di mismatch moderato, ma con un impatto dell'IA generalmente elevato. Il gruppo 5 (Professioni qualificate nelle attività commerciali e nei servizi) mostra una distribuzione eterogenea, con alcune professioni fortemente esposte all'IA, ma con un mismatch moderato, altre scarsamente esposte all'IA e con un mismatch basso.

Le professioni dei gruppi 6 (Artigiani, operai specializzati e agricoltori), 7 (Conduttori di impianti, operai di macchinari fissi e mobili e conducenti di veicoli) e 8 (Professioni non qualificate) sono situate tra il terzo e il quarto quadrante con valori di impatto dell'IA negativi e una contenuta variabilità per quanto riguarda il mismatch. Questo potrebbe riflettere la difficoltà di trovare una corrispondenza tra le competenze disponibili e quelle richieste, in settori caratterizzati da un potenziale di alta sostituibilità e bassi livelli di specializzazione.

Andando più in profondità nell'analisi dei profili CP2011 4 digit, emerge che l'8121 (Uscieri e professioni assimilate) presenta il livello di mismatch più elevato, pur registrando un impatto dell'IA leggermente negativo. Nonostante ciò, il livello di impatto dell'IA risulta nettamente superiore rispetto agli altri profili appartenenti al grande gruppo 8. Per quanto riguarda i profili maggiormente esposti all'impatto dell'Intelligenza artificiale, spiccano[4] il 1227 (Direttori e dirigenti generali di banche, assicurazioni, agenzie immobiliari e di intermediazione finanziaria), il 1231 (Direttori e dirigenti della finanza ed amministrazione), il 4311 (Addetti alla gestione degli acquisti) e il 3113 (Tecnici statistici). Al contrario, i professionisti del gruppo 6142 (Pulitori di facciate) risultano essere quelli con il mismatch più basso e tra i meno esposti all'impatto dell'IA.

Complessivamente numerose professioni appartengono al quadrante in alto a destra caratterizzato da alta esposizione all'IA e da elevato mismatch. Queste sono le professioni maggiormente 'interessanti' dato che sono quelle dove la IA potrà maggiormente aiutare a colmare gli squilibri del mercato del lavoro. Focalizzandoci su queste professioni la figura 3 ne mostra la distribuzione settoriale.

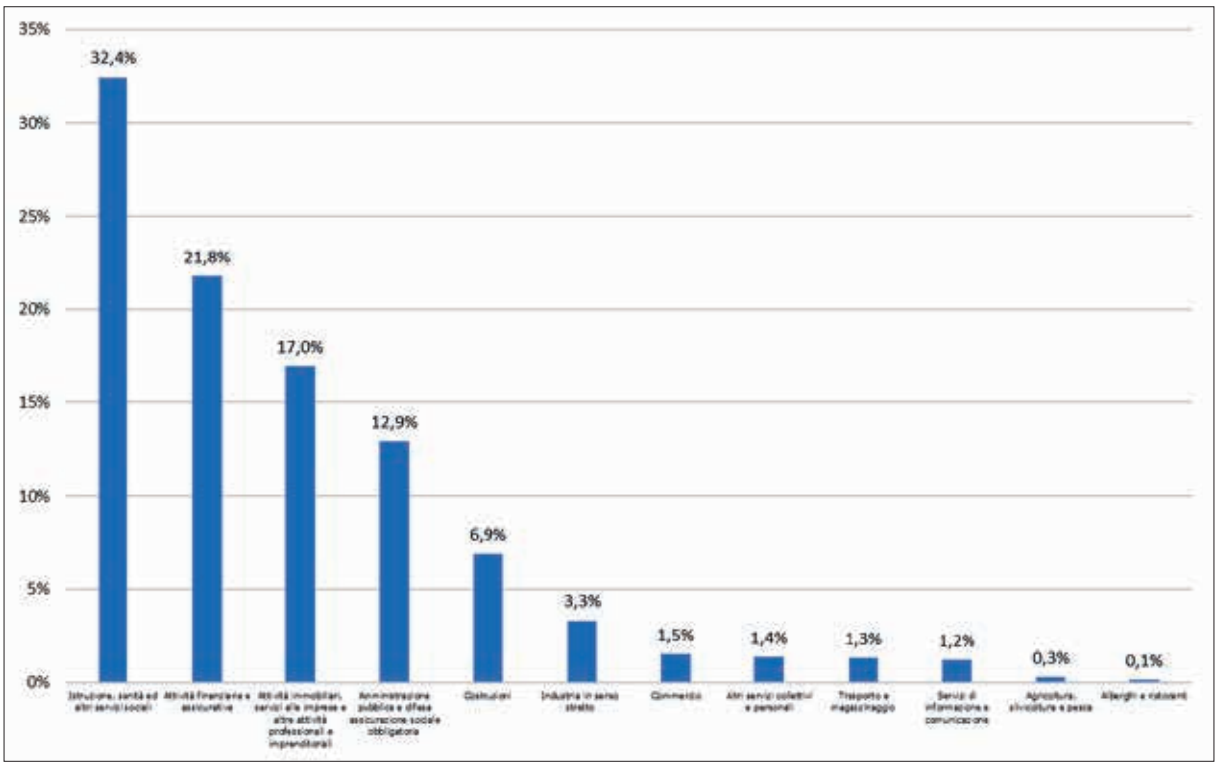
In generale, tali professioni coinvolgono circa 1,15 milioni di lavoratori, distribuiti in 12 macrosettori. Dall'analisi emerge che il settore con la maggiore concentrazione di questi profili è quello dell'istruzione, sanità e altri servizi sociali, che raccoglie il 32,4% del totale. Segue il comparto delle attività finanziarie e assicurative con il 21,8%, indicando una forte esposizione delle professioni altamente qualificate di questo settore agli effetti del disallineamento delle competenze e della tecnologia. Anche le attività immobiliari, i servizi alle imprese e altre attività professionali e imprenditoriali sono significativamente coinvolti (17%), suggerendo un'elevata incidenza di lavori con competenze soggette a trasformazioni tecnologiche e organizzative. L'amministrazione pubblica e la difesa, con il 12,9%, rappresentano un altro settore rilevante, mentre le costruzioni (6,9%) e l'industria in senso stretto (3,3%) mostrano una minore, ma comunque non trascurabile, presenza di lavoratori nelle professioni selezionate. I restanti settori (commercio, servizi collettivi e personali, trasporto e magazzinaggio, servizi di informazione e comunicazione, settore agricolo e quello del turismo) presentano quote più contenute, oscillando tra l'1,5% e lo 0,1%. Questi dati suggeriscono che l'impatto dell'IA e il mismatch delle competenze sono fenomeni particolarmente rilevanti nei settori a maggiore intensità di conoscenza e servizi, dove la trasformazione tecnologica e l'automazione stanno ridisegnando più rapidamente le richieste di competenze.

Per concludere, tramite un'analisi di regressione, è possibile catturare la relazione tra domanda

3 La dimensione dei punti corrisponde al numero di occupati per profilo professionale ISCO-08 2 digit (2023).

4 Sono stati inseriti i profili con un valore di impatto dell'IA superiore a 1,5.

Figura 3. Distribuzione settoriale delle professioni ad alta domanda insoddisfatta e alto impatto dell'Intelligenza artificiale in Italia



Fonte: elaborazioni Randstad Research su dati Cedefop, Rilevazione sulle Forze di lavoro (Istat) e Felten *et al.* (2021)

Tabella 3. Regressione dell'indice di esposizione all'IA sull'indicatore di mismatch per livello di occupazione

Numero di osservazioni	450			
F(2,447)	101,19			
Prob>F	0,0000			
R-squared	0,4649			
Root MSE	0,76168			
expIA	Coefficiente	Standard error	t	P>|t|
lmismatch	0,412011	0,1617588	2,55	0,011
skill	-1,162266	0,0617519	-18,82	0,000
Costante	1,757241	0,1029012	17,08	0,000

Fonte: elaborazioni Randstad Research su dati Cedefop, Rilevazione sulle Forze di lavoro (Istat) e Felten *et. al* (2021)

insoddisfatta ed esposizione all'IA. In termini analitici, definendo con *i* la professione ISCO-08 4 digit, *expIA* il grado di esposizione all'IA, *lmismatch* la misura di intensità di domanda insoddisfatta e con *skill*[5] una variabile discreta che identifica il livello di intensità o il livello di qualifica del lavoro[6]:

5 La variabile *skill* assume valore 1 per le professioni 'high skill', valore 2 per le professioni 'medium skill' e valore 3 per le professioni 'low skill'.

6 Le professioni sono così catalogate. I grandi gruppi ISCO-08 1, 2 e 3 costituiscono il gruppo delle professioni 'high skill'; i grandi gruppi 4, 5,6 e 7 costituiscono il gruppo delle professioni 'medium skill'; i grandi gruppi 8 e 9 costituiscono il gruppo delle professioni 'high skill'.

$$expIA_i = \alpha + \beta Imismatch_i + skill_i\,\gamma' + \epsilon_i$$

La tabella 3 mostra i risultati della stima del modello di regressione. L'analisi di correlazione mostra che sussiste una relazione positiva e statisticamente significativa tra esposizione all'IA e domanda insoddisfatta, indipendentemente dal livello di qualifica.

Conclusioni e discussione

L'analisi presentata ha delineato un quadro complesso delle dinamiche del mercato del lavoro italiano nel prossimo decennio, caratterizzato da un lato dalle sfide poste dall'invecchiamento della popolazione e dal fabbisogno occupazionale, e dall'altro dalle opportunità e dai rischi derivanti dalla crescente diffusione dell'Intelligenza artificiale.

I risultati evidenziano come l'IA stia già trasformando profondamente diversi settori e professioni, con un impatto significativo sia in termini di sostituzione di alcune mansioni sia di potenziamento di altre. La letteratura sul tema è molto varia e insiste molto sui possibili esiti distruttivi di posti di lavoro delle nuove tecnologie, arrivando a promuovere politiche di "mitigazione degli effetti dell'IA sul mercato del lavoro" (Ferri *et al.* 2024). La scelta di concentrarci sull'Indice AIOE di Felten *et al.* (2021) è stata dettata dal fatto che esso costituisce una delle misure più note nella letteratura. Ci ha permesso di condurre un'analisi approfondita e comparabile, ma siamo consapevoli che indici alternativi potrebbero catturare dimensioni diverse e complementari del fenomeno.

Per questo motivo, indichiamo come una naturale direzione per la ricerca futura la replica e l'estensione della nostra analisi utilizzando proprio questi indici alternativi.

Complessivamente le evidenze che emergono dal nostro lavoro suggeriscono che la sostituzione di posti di lavoro può essere auspicata in settori in cui vi sia domanda insoddisfatta, fermo restando che un aumento della produttività con l'adozione di soluzioni complementari sia auspicabile trasversalmente a tutte le professioni. La sostituzione di lavoratori sembra infatti la strategia maggiormente perseguita a livello individuale da parte delle imprese, per la tendenza tipica delle organizzazioni a standardizzare e automatizzare i processi (Acemoglu e Restrepo 2020). Le implicazioni di questa tendenza sono note e includono una riduzione del contributo del lavoro al valore aggiunto, e possono spiegare la riduzione della produttività, in particolare nei servizi, e come conseguenza una contrazione della quota salari riconosciuta al lavoro, sia direttamente per la minore produttività sia indirettamente per il ridotto potere contrattuale della parte lavoratrice. Tuttavia, in questa fase storica di invecchiamento della popolazione, di riduzione della popolazione attiva e della forza lavoro anche a fronte di un tasso di attività in crescita, si prospetta come una parziale soluzione alla domanda insoddisfatta.

Analizzando sotto questa lente il caso italiano, e partendo dalle evidenze presentate in Ferri *et al.* (2024), è interessante notare che per una parte delle professioni più esposte si prevede una domanda insoddisfatta piuttosto elevata. È il caso degli Addetti agli affari generali (CP2011 4.1.1.2), degli Addetti alle funzioni di segreteria (CP2011 4.1.1.1) e degli Agenti di commercio (CP2011 3.3.4.2), per i quali si prevede un'elevata domanda insoddisfatta, intorno al terzo quartile della distribuzione, con un fabbisogno insoddisfatto rispettivamente di 267.342 nel periodo 2024-2033 (25,9% in rapporto agli occupati del 2023), 58.458 (16,5%) e 77.131 (43,9%). La sostituzione delle mansioni svolte da queste professioni con strumenti IA può quindi sopperire alla mancanza di occupati. I cambiamenti organizzativi che ne derivano non sono in ogni caso per nulla ovvi, in quanto ogni organizzazione può scegliere o di spostare le mansioni su altre professioni, facendo di fatto scomparire il ruolo e provocando un cambiamento al margine estensivo, oppure di mantenere formalmente la professione ma mutarne i contenuti, ad esempio configurandola come supervisione di strumenti automatizzati. In altri casi la sostituzione, seppur teoricamente possibile, sembra essere poco probabile. Infatti, pur esistendo la possibilità tecnica di introdurre strumenti basati sull'IA, la difficoltà di reperimento è complessivamente ridotta e insufficiente a generare un incentivo a innovare a favore di tecnologie ad alta intensità di capitale. Questa eventualità ben si adatta al caso italiano: i salari reali si sono stabilmente ridotti negli ultimi anni (Evangelista e Pacelli 2025; OECD 2024).

La nostra analisi permette di isolare alcune professioni per le quali ci aspettiamo questi esiti. È il caso degli Addetti alla gestione dei magazzini (CP2011 4.3.1.2) e Tecnici programmatori (CP2011 3.1.2.1) per i quali la domanda insoddisfatta è

relativamente bassa e rispettivamente di -4.334 (1,5%) e -5.321 (2,9%).

Questa dinamica è ben descritta nel lavoro di Fenoaltea *et al.* (2024), che mostrano come molte professioni con ruoli decisionali o che si occupano della pianificazione strategica pur avendo un'elevata esposizione all'IA, non siano concretamente bersaglio delle startup a causa di barriere etiche, sociali o legate all'alta posta in gioco (*high-stakes*) che vanno oltre la semplice fattibilità tecnica.

Nel contesto di trasformazione in cui gli esiti dell'adozione dell'IA sono ancora in via di definizione, la formazione continua, sia con interventi di aggiornamento (*upskilling*) sia di ampliamento delle competenze (*reskilling*) assume un ruolo cruciale per consentire ai lavoratori di accogliere i cambiamenti tecnici e organizzativi in modo attivo (Morandini *et al.* 2023). Gli studi hanno rivelato che la complementarità tra IA e abilità umane è maggiore per professioni ad alta qualifica e alti redditi, mostrando indirettamente che il livello di istruzione è un fattore protettivo rispetto alla sostituibilità (Pizzinelli *et al.* 2023).

Investire in programmi di formazione mirati, che sviluppino le competenze necessarie per interagire con le nuove tecnologie e per svolgere mansioni a più alto valore aggiunto, è fondamentale per ridurre il mismatch occupazionale e per garantire una transizione equa e inclusiva. È inoltre importante notare che esposizione all'IA non implica la necessità di possedere competenze specialistiche, come il *machine learning* o il *natural language processing*, bensì di gestire questi nuovi strumenti e di saperli integrare proficuamente nei processi di lavoro (Green 2024). Sono quindi maggiormente richieste competenze quali la gestione generale dei progetti, finanza, amministrazione e compiti di segreteria. Inoltre, le competenze di comunicazione, *problem solving*, creatività e lavoro di squadra sono spesso richieste in relazione a ruoli che prevedono l'utilizzo di strumenti IA, domanda consolidata già nel periodo di iniziale adozione di queste tecnologie (Squicciarini e Nachtigall 2021). Il disegno di percorsi di formazione dovrà essere individualizzato in quanto il valore di ogni nuova competenza appresa dipende dall'insieme delle competenze già possedute dal lavoratore, in particolare se riguardanti l'IA (Stephany e Teuloff 2024).

Inoltre, l'innovazione digitale, se da un lato può rappresentare una sfida per l'occupazione in alcuni settori, dall'altro offre importanti opportunità di *retention* per i lavoratori. La possibilità di essere coinvolti in progetti innovativi, di utilizzare strumenti all'avanguardia e di acquisire nuove competenze può aumentare la motivazione e l'engagement dei dipendenti, riducendo il rischio di turnover e favorendo la creazione di un ambiente di lavoro stimolante e dinamico. Le aziende che sapranno investire nella formazione dei propri dipendenti e promuovere una cultura dell'innovazione continua saranno in grado di trattenere i talenti, di migliorare la produttività e di affrontare con successo le sfide del mercato del lavoro del futuro.

Bibliografia

Acemoglu D., Restrepo P. (2020), The wrong kind of AI? Artificial intelligence and the future of labour demand, *Cambridge Journal of Regions, Economy and Society*, 13, n.1, pp.25-35

Albora G., Diodato D., Fenoaltea E.M., Mazzilli D., Patelli A., Sbardella A., Sciarra C., Tacchella A., Zaccaria A. (2025), *Challenges and opportunities for the EU labour market from AI development*, JRC Working Paper n.JRC141782, Bruxelles, European Commission

Autor D.H. (2013), The "task approach" to labor markets: An overview, *Journal for Labour Market Research*, 46, n.3, pp.185-199

Eloundou T., Manning S., Mishkin P., Rock D. (2024), GPTs are GPTs: Labor market impact potential of LLMs, *Science*, 384, n.6702, pp.1306-1308

Evangelista R., Pacelli L. (2025), *Lavoro e salari in Italia: cambiamenti nell'occupazione, precarietà, impoverimento*, Roma, Carocci

Felten E., Raj M., Seamans R. (2021), Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses, *Strategic Management Journal*, 42, n.12, pp.2195-2217

Fenoaltea E.M., Mazzilli D., Patelli A., Sbardella A., Tacchella A., Zaccaria A., Trombetti M., Pietronero L. (2024), Follow the money: A startup-based measure of AI exposure across occupations, industries and regions, *arXiv* <arXiv:2412.04924>

Ferri V., Porcelli R., Fenoaltea E.M. (2024), *Lavoro e Intelligenza Artificiale in Italia: tra opportunità e rischio di sostituzione*, Inapp Working Paper n.125, Roma, Inapp

Green A. (2024), *Artificial intelligence and the changing demand for skills in the labour market*, OECD Artificial Intelligence Papers n.14, Paris, OECD Publishing

Istat (2024), *Rapporto Annuale 2024. La situazione del Paese*, Roma, Istat

Morandini S., Fraboni F., De Angelis M., Puzzo G., Giusino D., Pietrantoni L. (2023), The impact of artificial intelligence on workers'skills: Upskilling and reskilling in organisations, *Informing Science: The International Journal of an Emerging Transdiscipline*, 26, pp.39-68

Pizzinelli C., Panton A., Mendes Tavares M., Cazzaniga M., Li L. (2023), *Labor market exposure to AI: Cross-country differences and distributional implications*, IMF Working Paper n.23/216, Washington, International Monetary Fund

Squicciarini M., Nachtigall H. (2021), *Demand for AI skills in jobs: Evidence from online job postings*, OECD Science, Technology and Industry Working Papers n.2021/03, Paris, OECD Publishing

Simons W., Turrini A., Vivian L. (2024), *Artificial intelligence: Economic impact, opportunities, challenges, implications for policy*, Discussion Paper n.210, Luxembourg, Pubblication Office of the European Union

Stephany F., Teutloff O. (2024), What is the price of a skill? The value of complementarity, *Research Policy*, 53, n.1 <https://doi.org/10.1016/j.respol.2023.104898>

Giovanni Armillotta
giovanni.armillotta@randstad.it
È ricercatore statistico presso Randstad Research Institute dal 2022. Ha conseguito una Laurea magistrale in Statistica e Metodi per l'economia e la finanza presso l'Università degli Studi di Bari Aldo Moro. Si occupa principalmente di ricerche sul mercato del lavoro nazionale e internazionale, con un focus sulle competenze dei profili professionali e la loro evoluzione negli anni.

Emilio Colombo
emilio.colombo@unicatt.it
Ha conseguito il Master of Science e il Dottorato di ricerca presso l'Università di Southampton. È coordinatore del Comitato scientifico di Randstad Research Institute e Professore Ordinario di Politica economica presso l'Università Cattolica del Sacro Cuore. È autore di numerosi articoli nelle principali riviste internazionali, di numerosi saggi e libri. Ha insegnato in diverse università svolgendo seminari in varie sedi internazionali. Ha partecipato a numerosi progetti internazionali legati all'analisi delle skill e del mercato del lavoro. I suoi interessi di ricerca sono nell'economia internazionale, nell'economia del lavoro e nell'economia applicata.

Martina Gnudi
martina.gnudi@randstad.it
È ricercatrice statistica presso Randstad Research Institute dal 2019. Ha conseguito una Laurea magistrale in Statistica, economia e impresa con un focus su Marketing e ricerche di mercato presso l'Università di Bologna.

Francesca Lettieri
francesca.lettieri@randstad.it
È ricercatrice qualitativa e editor presso Randstad Research Institute. Si occupa dello studio dei megatrend, in particolare sostenibilità, cambiamenti demografici e digitalizzazione e del loro impatto nel plasmare domanda e offerta di competenze. Con riferimento a quest'ultimo tema sta collaborando ad una serie di approfondimenti dedicati al tema dell'Intelligenza artificiale.

Federica Romano
federica.romano@randstad.it
È coordinatrice di Randstad Research Institute. Laureata in Giurisprudenza presso la Luiss Guido Carli, ha conseguito un Dottorato di ricerca in Formazione della persona e mercato del lavoro presso l'Università degli Studi di Bergamo. È stata Research Fellow presso Adapt, Centro Studi nell'ambito delle Relazioni industriali e di lavoro e Visiting Research Fellow presso Eurofound, EU Agency for the improvement of living and working conditions.

Francesco Trentini
francesco.trentini@unimib.it
È un economista, Assegnista di ricerca presso l'Università degli Studi di Milano-Bicocca, collabora con il Centro di ricerca interuniversitario sui Servizi pubblici (CRISP) e con il LABORatorio R. Revelli. È coordinatore della ricerca presso Randstad Research Institute. I suoi interessi di ricerca riguardano l'economia del lavoro, con particolare attenzione all'evoluzione delle professioni, alla performance del mercato del lavoro e al Terzo settore. Svolge primariamente attività di ricerca quantitativa, in particolare con dati amministrativi e dati web.

Saggi

Invecchiamento della forza lavoro in Italia

Analisi dei cambiamenti strutturali e delle dinamiche intergenerazionali nel mercato del lavoro

Giuseppe Venere
Università degli studi di Bari Aldo Moro

Giorgia Panico
Università del Salento

Il mercato del lavoro italiano è oggetto di una profonda trasformazione: la quota di lavoratori over 55 cresce costantemente, mentre i giovani faticano a trovare spazio. Sebbene il dibattito si concentri su una presunta competizione intergenerazionale, le evidenze empiriche non ne confermano l'esistenza. La crescita dell'occupazione senior non rappresenta, difatti, un ostacolo all'ingresso dei giovani nel mercato del lavoro, ma sollecita l'adozione di politiche integrate volte a potenziare le opportunità giovanili, ridurre le disparità territoriali e implementare pratiche di *age management* in azienda.

The Italian labour market is undergoing a significant structural change: the proportion of workers aged 55 and above is steadily rising, while young people encounter persistent obstacles to labour market entry. Although public debate frequently focuses on purported intergenerational competition, empirical data does not support this claim. Rather, the expansion of senior employment highlights the need for integrated policy frameworks to enhance youth opportunities, address regional disparities, and embed age management practices in the workplace.

DOI: 10.53223/Sinappsi_2025-02-16

Citazione
Venere G., Panico G. (2025), Invecchiamento della forza lavoro in Italia. Analisi dei cambiamenti strutturali e delle dinamiche intergenerazionali nel mercato del lavoro, *Sinappsi*, XV, n.2, pp.243-256

Parole chiave
Invecchiamento attivo
Occupazione giovanile
Organizzazione del lavoro

Keywords
Active ageing
Youth employment
Work organization

1. La nuova sfida demografica: il lavoro

L'invecchiamento della popolazione attiva in Italia rappresenta una sfida complessa, con implicazioni significative per il mercato del lavoro e l'economia nazionale. Negli ultimi vent'anni, la quota di lavoratori over 55 è cresciuta considerevolmente (Vignoli e Tomassini 2023; Istat 2024; Rosina 2024), con aumenti particolarmente marcati in alcune regioni, dove fattori socioeconomici e normativi spingono verso un'uscita dal lavoro più tardiva (Charness e Villeval 2009; Belloni e Maccheroni 2013; Maccheroni *et al.* 2025).

La crescente presenza di lavoratori senior solleva interrogativi sulla potenziale competizione intergenerazionale, soprattutto in contesti caratterizzati da una domanda di lavoro stagnante o soggetta a contrazioni cicliche. In alcune aree del Paese, in particolare nel Mezzogiorno, si potrebbe ipotizzare che, a causa delle disparità economiche tra regioni, i giovani percepiscano la presenza di lavoratori senior come un ostacolo al loro ingresso nel mercato del lavoro (Cazzola 2004; Socci 2015; Rudolph e Zacher 2015; Della Porta *et al.* 2021). Tuttavia, la convivenza tra

fasce d'età diverse può generare sinergie: il trasferimento di competenze ed esperienza si è dimostrato, difatti, un'opportunità (Ali e French 2019; Firzly *et al.* 2021). Per gestire efficacemente questa complessità, è fondamentale ridurre le disuguaglianze generazionali e territoriali, tenendo conto della rapida digitalizzazione e delle riorganizzazioni aziendali che stanno ridefinendo il lavoro a livello globale. In questo contesto, l'innovazione tecnologica diventa una leva fondamentale per l'aggiornamento continuo delle competenze e il sostegno alla longevità lavorativa, obiettivo che esige un approccio integrato che bilanci le esigenze delle diverse fasce d'età con le peculiarità dei contesti produttivi (Billari e Tomassini 2024; Inapp *et al.* 2025).

Le teorie demografiche e sociali sui mutamenti della popolazione attiva suggeriscono che l'allungamento della vita lavorativa e il ridotto ricambio generazionale possano compromettere produttività e competitività (Burke *et al.* 2013; Behaghel *et al.* 2014). Di conseguenza, strategie che incentivano l'aggiornamento continuo delle competenze, il passaggio intergenerazionale delle conoscenze e l'adozione di pratiche di age management diventano cruciali, specialmente se accompagnate da riforme previdenziali mirate a sostenere la longevità lavorativa (Ripamonti *et al.* 2021; Marcus *et al.* 2024). Dal punto di vista territoriale, le aree più dinamiche integrano meglio i senior senza penalizzare i giovani, mentre regioni meno sviluppate alimentano la percezione di competizione (Marcaletti 2013; Socci 2015; Della Porta *et al.* 2021). Perciò, è necessario adottare politiche flessibili e calibrate sulle specificità locali, coniugando sostegno all'occupazione senior e investimenti per i giovani (Struffolino e Filandri 2021). Pur non essendo il focus principale, l'utilizzo di variabili territoriali permetterà di contestualizzare le dinamiche intergenerazionali e spiegare le differenze strutturali tra le macroaree italiane (Benassi *et al.* 2021). L'obiettivo di questo studio è indagare se e in che misura la crescente presenza di lavoratori over 55 possa generare una competizione con i giovani o, al contrario, favorire un modello di sviluppo fondato sul trasferimento intergenerazionale di conoscenze ed esperienze (Rudolph e Zacher 2015). Le principali domande di ricerca sono: *esiste una relazione tra il Tasso di occupazione senior (TOS, 55-64 anni) e il Tasso di disoccupazione giovani (TDG, 18-29 anni) in Italia, tenendo conto delle variazioni territoriali e temporali? Quali interventi normativi hanno influenzato la struttura e le dinamiche del mercato del lavoro italiano, e in quale contesto economico sono stati introdotti?*

L'analisi, basata sui dati dell'Istituto nazionale di statistica (Istat), utilizza tecniche di correlazione annuale, attraverso lo Spearman Test e la correlazione di Pearson, per fornire un quadro dettagliato delle dinamiche intergenerazionali nel contesto italiano. Nella seconda parte del contributo si illustrano gli interventi normativi che hanno interessato il mercato del lavoro italiano negli ultimi decenni all'interno del più ampio contesto economico, integrando tale analisi con una rassegna della letteratura sociologica ed economica di riferimento.

2. Italia: incremento della popolazione senior

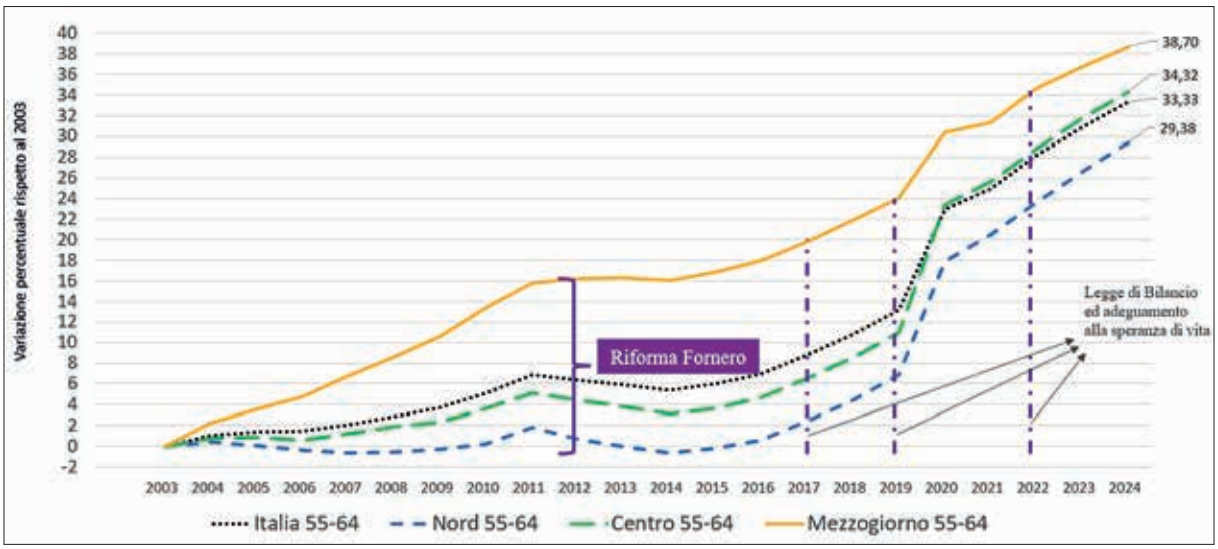
Secondo i dati aggregati a livello nazionale, la popolazione di età compresa fra 55 e 64 anni ha progressivamente aumentato la propria incidenza all'interno del mercato del lavoro (Benassi *et al.* 2021; Istat 2016; 2019; 2024). L'innalzamento dell'età pensionabile, le trasformazioni socio-culturali che rendono più frequente il proseguimento dell'attività lavorativa in età avanzata e la necessità individuale di consolidare i requisiti contributivi hanno determinato, nel loro complesso, una crescita costante del Tasso di occupazione senior (TOS). Questa dinamica non risulta omogenea sull'intero territorio, ma risente delle peculiarità produttive e delle opportunità di impiego offerte nelle diverse macroaree.

Come già accennato, l'aumento dell'occupazione nella fascia senior rispecchia in parte anche l'espansione della popolazione stessa in questa classe d'età. Prendendo in esame un periodo di circa vent'anni, si osserva infatti come la coorte 55-64 sia cresciuta in Italia e, con intensità diverse, in ciascuna macroarea rispetto al periodo di riferimento. Per studiare tale fenomeno si è ricorso a una misura di variazione percentuale con base 2003:

$$V\%^{area}_{55-64(t)} = \left(\frac{P^{area}_{55-64(t)}}{P^{area}_{55-64(2003)}} * 100\right) - 100 \;\; \forall_t \in [2003,\ldots,2004]$$

dove la popolazione $P^{area}_{55-64(t)}$ rappresenta gli individui nella fascia 55-64 anni residenti in una determinata area geografica (Italia nel suo complesso, Nord, Centro e Mezzogiorno) all'anno t; il 2003 costituisce l'anno base.

Figura 1. Variazione percentuale della popolazione in fascia 55-64 anni in Italia e per macroaree (Nord, Centro e Mezzogiorno) dal 2003 al 2024 (base 2003=0)



Fonte: elaborazioni su dati Istat, 2024

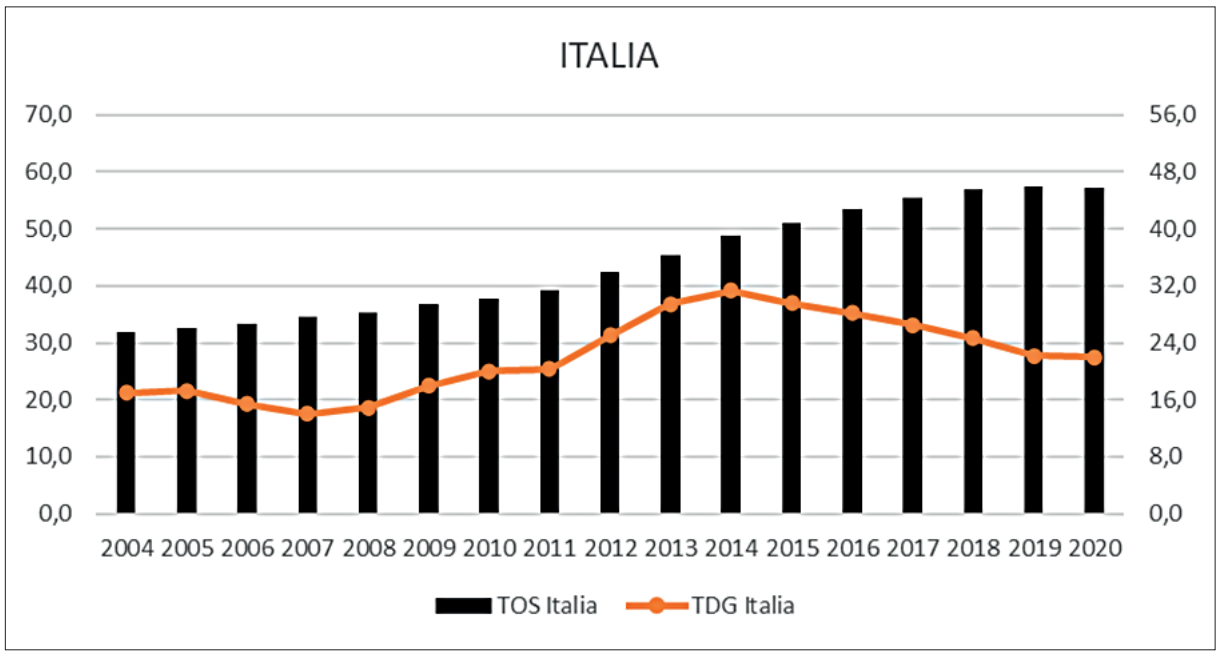
Dall'osservazione della figura 1 emergono quattro linee distinte: una per l'andamento medio italiano e una per ciascuna delle principali macroaree. Sull'asse delle ordinate è riportata la variazione percentuale rispetto al 2003. Nei primissimi anni della serie (2003-2005), i valori si collocano prossimi allo zero, poiché la popolazione senior non si discosta significativamente dalla base di riferimento. A partire dal 2006-2007, l'incremento diventa progressivamente più evidente e si accentua ulteriormente in corrispondenza delle principali riforme normative – evidenziate dalle linee verticali – come la Riforma Fornero e le successive Leggi di Bilancio. Le curve mostrano dunque un graduale innalzamento, con qualche rallentamento nei periodi di minore dinamismo economico, ma mantenendo nell'insieme un trend crescente. Nel periodo più recente, dal 2017 fino al 2024, si registrano i livelli più elevati di variazione percentuale. Il Mezzogiorno raggiunge il punto più alto della curva, arrivando al 38,7% di crescita rispetto al 2003, mentre il Nord supera il 34% e l'Italia nel suo complesso si attesta al 33,3%. Il Centro rimane leggermente al di sotto, intorno al 29,4%, evidenziando un andamento costante ma meno pronunciato rispetto alle altre aree geografiche. Le differenze fra le varie curve sottolineano come, pur essendo un fenomeno generale, l'incremento della popolazione nella fascia 55-64 risulti particolarmente significativo in alcuni contesti regionali.

L'espansione osservata può essere stata favorita anche dalle particolari dinamiche demografiche che hanno ridotto la base della piramide dell'età italiana, riconducibili sia al calo della fecondità, sia alla mobilità di fasce giovanili verso altri territori (Billari 2023; Vignoli e Tomassini 2023). Come precedentemente menzionato, risulta interessante osservare come l'incremento della popolazione in età senior sia avvenuto in concomitanza con significative modifiche legislative in ambito lavorativo e pensionistico.

3. Confronto tra Tasso di occupazione senior (TOS) e Tasso di disoccupazione giovani (TDG) a livello nazionale

L'analisi della figura 2 accosta l'evoluzione del Tasso di occupazione senior (TOS) a quella del Tasso di disoccupazione giovani (TDG) per il periodo 2004-2020. L'obiettivo è osservare in che modo la crescita dell'occupazione tra gli over 55 (asse sinistro) si collochi rispetto all'andamento della disoccupazione giovanile nella fascia 18-29 (asse destro).

Figura 2. Andamento del Tasso di occupazione senior (TOS, asse sinistro) e del Tasso di disoccupazione giovani (TDG, asse destro) in Italia, 2004-2020



Fonte: elaborazioni su dati Istat, 2024

$$TOS = \frac{Occupati_{55-64}}{Popolazione_{55-64}} * 100$$

$$TDG = \frac{Disoccupati_{18-29}}{Disoccupati_{18-29} + Occupati_{18-29}} * 100$$

L'incremento del TOS appare graduale, con lievi rallentamenti in prossimità dei principali momenti di instabilità economica; il TDG, invece, risulta maggiormente soggetto a variazioni congiunturali, registrando picchi durante i periodi di recessione economica.

Come si osserva in figura 2, a partire dal 2014 l'incremento continuo del TOS coincide con una graduale riduzione del Tasso di disoccupazione giovanile, che raggiunge il 22% verso la fine del periodo osservato. Questa dinamica indica che, pur in presenza di un TOS in costante aumento, le opportunità per i giovani non subiscono un effetto di sostituzione lineare in alcuni intervalli temporali, come i dati potevano inizialmente far ipotizzare: a partire dal 2014 in poi si rileva, infatti, un cambiamento di tendenza.

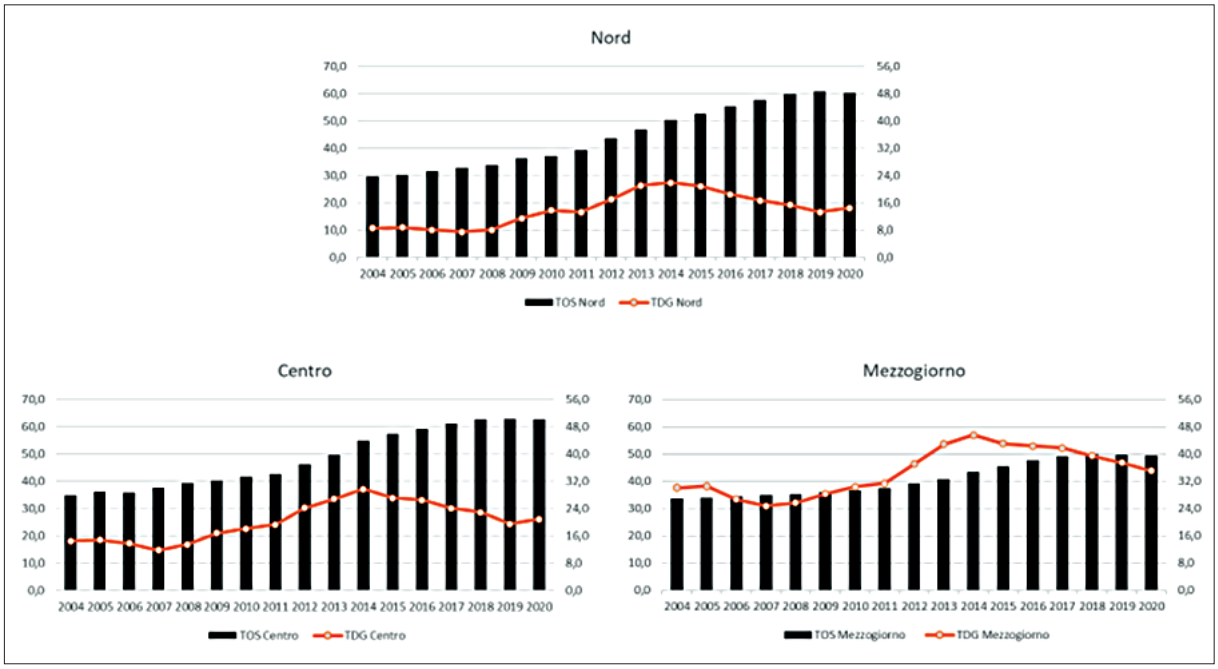
Passando all'osservazione delle tre macroaree, in cui i medesimi dati sono suddivisi per Nord, Centro e Mezzogiorno d'Italia (figura 3) emerge quanto di seguito descritto.

Nel Nord Italia, il TOS registra un aumento più accentuato rispetto alla media nazionale, raggiungendo o superando il 60% verso la fine del periodo. La disoccupazione giovanile segue un andamento altalenante, ma con valori contenuti rispetto alle altre aree geografiche, generalmente vicini al 20-22%, nei momenti più critici. La curva del TDG presenta picchi in corrispondenza degli stessi anni di crisi economica già rilevati a livello nazionale (2013-2014), ma questa risulta meno elevata se confrontata con le altre ripartizioni territoriali.

Nel Centro Italia, il TOS supera lievemente quello registrato nel Nord, attestandosi al di sopra della media nazionale in alcuni anni-chiave. La linea del TDG, tuttavia, evidenzia variazioni più accentuate con oscillazioni; si osserva, in particolare, un picco superiore al 30% nell'arco temporale 2012-2017, in concomitanza con la fase di maggiore criticità economica del periodo considerato.

Nel Mezzogiorno d'Italia, il grafico mostra un TOS che parte da valori piuttosto bassi rispetto alle altre ripartizioni territoriali, prossimi al 35% nei primi anni della serie, per poi aumentare gradualmente sino a stabilizzarsi attorno al 49% a partire dal 2016. La curva del TDG si distanzia nettamente da quella del TOS, superando il

Figura 3. Andamento del Tasso di occupazione senior (TOS, asse sinistro) e del Tasso di disoccupazione giovani (TDG, asse destro) nelle macroaree Nord, Centro e Mezzogiorno, 2004-2020



Fonte: elaborazioni su dati Istat, 2024

45% in diversi periodi ed evidenziando tassi di disoccupazione giovanile molto più elevati rispetto alle altre macroaree e alla media nazionale. Questi livelli di disoccupazione giovanile, associati a bassi tassi di occupazione senior, sottolineano la fragilità strutturale del mercato del lavoro meridionale, dove i giovani di età compresa fra 18-29 anni risultano particolarmente penalizzati nelle fasi congiunturali più critiche.

L'analisi dei dati sul mercato del lavoro italiano attraverso il TOS e il TDG per macroaree geografiche si inserisce, quindi, nel solco del tradizionale divario economico e occupazionale tra Nord e Sud, che rappresenta una caratteristica strutturale del Paese (Reyneri 2017; Orientale Caputo 2021; Viesti 2021). La frattura fra macroaree emerge chiaramente: a destare le maggiori preoccupazioni è la condizione occupazionale giovanile, particolarmente critica nel Mezzogiorno, le cui cause sono da attribuire a molteplici fattori, quali i ritardi nella transizione all'età adulta (scuola-lavoro) (Garcìa-Pereiro *et al.* 2024), i bassi livelli di istruzione – con una quota elevata di giovani meridionali in possesso della sola licenza media – e la persistente carenza di infrastrutture sociali e di politiche attive per l'occupazione giovanile (Orientale Caputo 2021).

4. Quando l'esperienza incontra il futuro: la relazione tra TOS e TDG

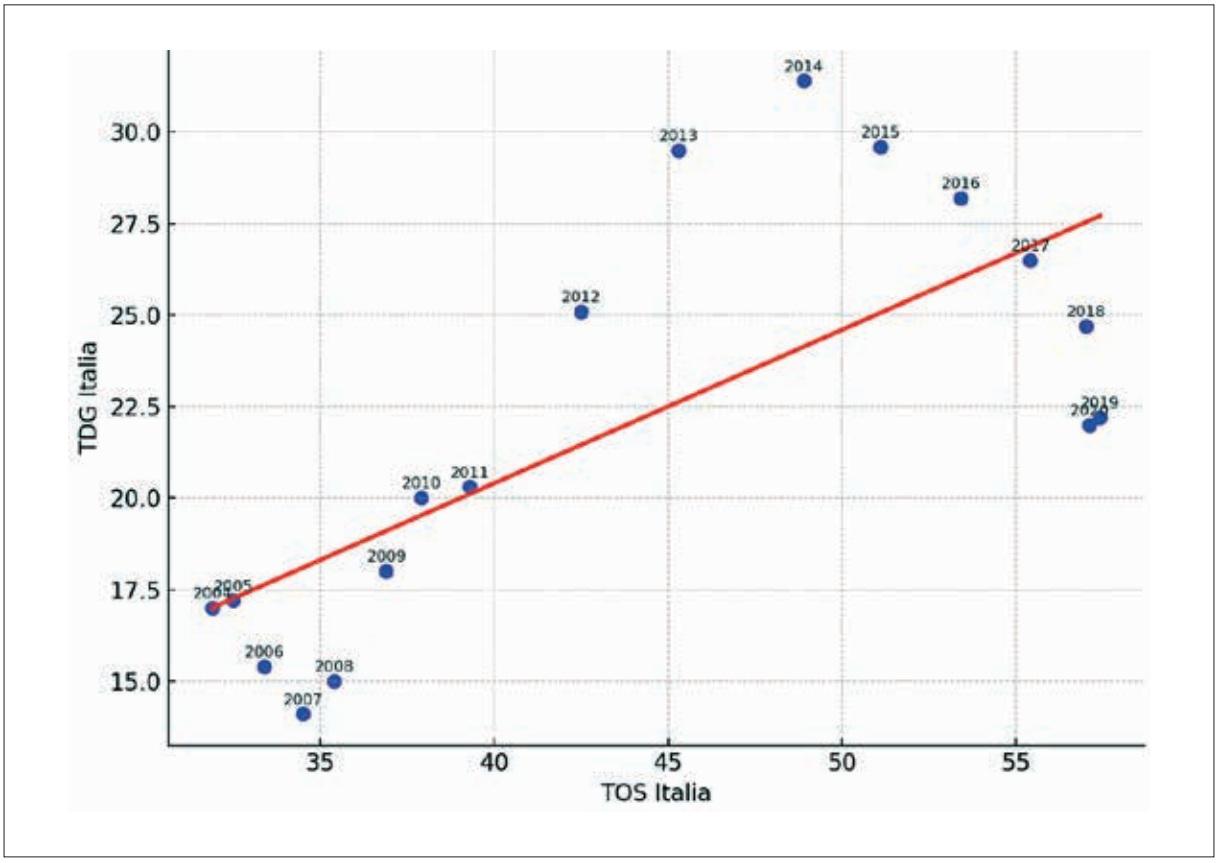
L'analisi (figura 4) della correlazione tra il Tasso di occupazione senior (TOS) e del Tasso di disoccupazione giovani (TDG) tra il 2004 e il 2020 consente di osservare una dinamica che, a una prima lettura, sembrerebbe suggerire una relazione tra le due variabili.

Nei primi anni considerati, il TOS – rappresentato in ascissa – segue un trend in crescita, passando dal 31,9% nel 2004 al 48,9% nel 2014. Parallelamente, il TDG – rappresentato in ordinata – mostra anch'esso un incremento, raggiungendo nello stesso anno il 31,4%. In questa prima fase, le due variabili sembrano muoversi nella stessa direzione: all'aumento dell'occupazione tra i lavoratori senior si associa una crescita della disoccupazione giovanile.

Dopo il 2014, invece, la direzione di tendenza delle due variabili si inverte (figura 4): il TOS continua ad aumentare fino a superare il 57% nel 2018, mentre il TDG diminuisce gradualmente fino al 22,2% nel 2019, assestandosi al 22% nel 2020.

Per verificare ulteriormente la correlazione, è stato eseguito il test di Correlazione di Spearman,

Figura 4. Coefficiente di correlazione di Spearman (rho di Spearman) tra il Tasso di disoccupazione giovani e il Tasso di occupazione senior in Italia, 2004-2020



Fonte: elaborazioni su dati Istat, 2024

che consente di misurare la correlazione, ossia il grado di associazione, monotona tra due variabili ordinali o quantitative. Il fine è comprendere se esista una relazione crescente o decrescente tra due serie di dati, senza richiedere che tale relazione sia lineare, né che i dati siano distribuiti normalmente (a differenza del test di Pearson). L'analisi ha impiegato la seguente formula:

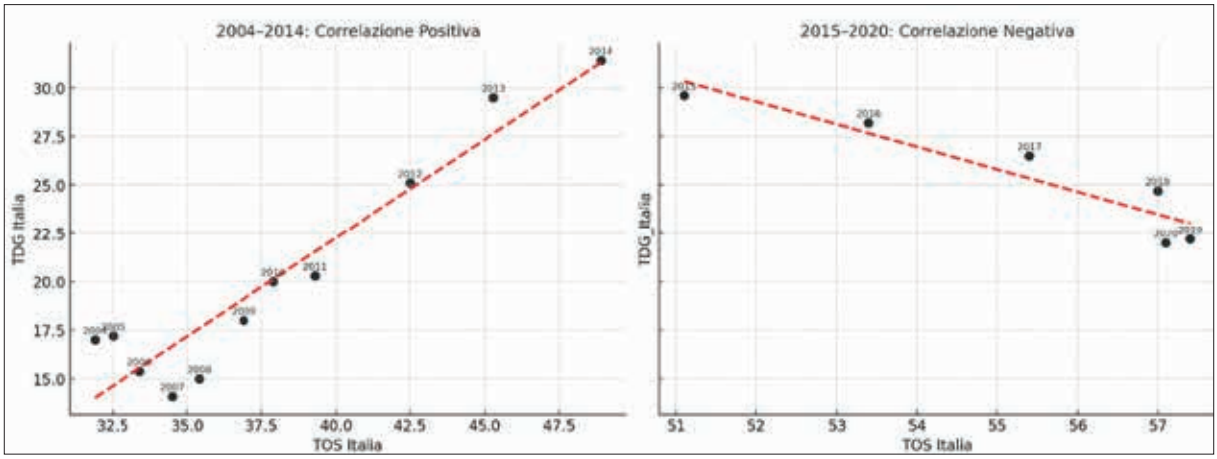
$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

dove d_i rappresenta la differenza tra i ranghi della coppia i-esima, mentre n il numero di osservazioni (una per ogni anno).

L'analisi condotta sull'intero periodo 2004-2020 (figura 4) restituisce un coefficiente di correlazione di Spearman pari a rho = 0,708, evidenziando una forte associazione monotona positiva: all'aumentare del TOS Italia, tende ad aumentare anche il TDG Italia. Il *p-value* di 0,0015 conferma la significatività statistica del risultato, suggerendo come tale relazione sia difficilmente attribuibile al caso. È importante sottolineare che la correlazione osservata tra le due variabili non implica necessariamente un rapporto di causalità. Il trend rilevato, infatti, potrebbe riflettere l'influenza di fattori esterni non considerati nell'analisi, quali cicli economici, modifiche alla legislazione del mercato del lavoro, dinamiche demografiche o variazioni nella domanda e offerta di lavoro.

Per facilitare una lettura più chiara del fenomeno, si è operata una suddivisione del periodo considerato in due sotto-intervalli. Nel primo sotto-periodo (2004-2014, figura 5), la correlazione tra TDG e TOS risulta fortemente positiva (rho = 0,836; p = 0,0013), indicando che nei primi anni dell'intervallo entrambi gli indicatori tendono ad aumentare simultaneamente: all'aumentare del TOS, cresce anche il TDG. Nel secondo periodo (2015-2020, figura 5), la correlazione si inverte: il coefficiente di Spearman assume un valore negativo molto

Figura 5. Coefficiente di correlazione di Spearman (rho di Spearman) tra il Tasso di disoccupazione giovani e il Tasso di occupazione senior in Italia, 2004-2014 e 2015-2020



Fonte: elaborazioni su dati Istat, 2024

marcato (rho = -0,943; p = 0,0048), suggerendo che, in questa fase, l'incremento dell'occupazione tra gli over 55 si associa a una diminuzione della disoccupazione giovanile. In entrambi i casi, è importante sottolineare che il test di Spearman non implica una relazione causale tra le due variabili, ma segnala esclusivamente un'associazione monotona – crescente nel primo periodo, decrescente nel secondo – ossia un andamento congiunto regolare, sebbene non necessariamente lineare o perfetto.

L'andamento, in effetti, potrebbe riflettere un miglioramento complessivo del mercato del lavoro, legato ad esempio a una fase di espansione economica o a interventi normativi che hanno favorito l'occupazione in generale. In tal senso, non si può affermare che l'aumento dell'occupazione senior comporti automaticamente un miglioramento della condizione giovanile.

Questa inversione di tendenza solleva alcuni quesiti importanti, sintetizzati nelle seguenti domande: quali fattori hanno determinato il cambiamento osservato a partire dal 2014? È plausibile che l'uscita dalla fase più critica della crisi economica, le riforme del mercato del lavoro e le nuove condizioni occupazionali abbiano contribuito a modificare il rapporto tra le due fasce di età, riducendo il potenziale conflitto generazionale? Quali sono stati i principali elementi di questa transizione e in che misura il mercato del lavoro italiano è riuscito a adattarsi a questi cambiamenti, garantendo al contempo l'accesso dei giovani al mercato del lavoro? A queste domande si tenterà di rispondere nei paragrafi successivi (cfr. paragrafo 5).

5. La relazione tra i due tassi: la non correlazione

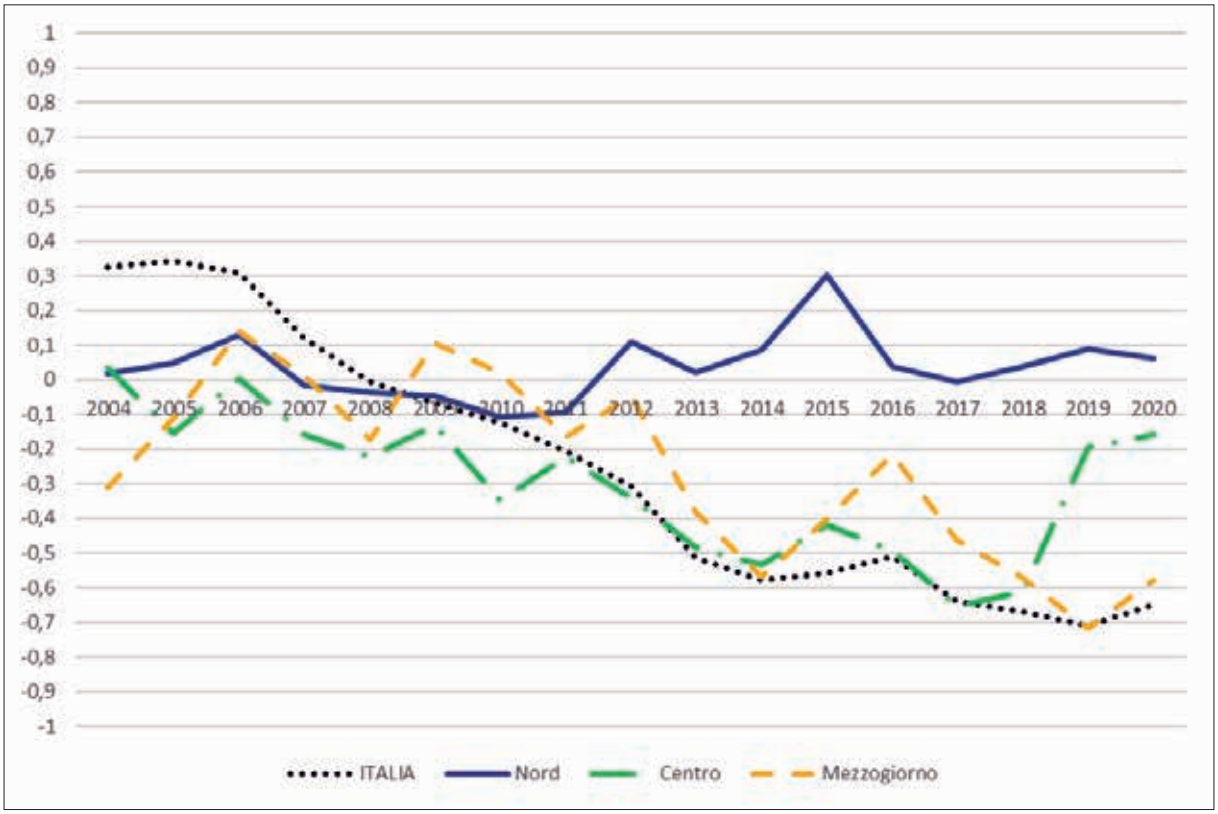
Si procede ora a stimare la relazione reciproca e di tipo lineare tra le due variabili, considerando le diverse aree territoriali, mediante la correlazione di Pearson calcolata annualmente nel periodo 2004-2020.

$$r\ Pearson\ (TOS, TDG)\ \forall_t \in [2004, \ldots, 2020]$$

Tale calcolo è ripetuto 'per ogni anno', prendendo in considerazione, a seconda dei casi, i dati delle diverse province aggregando per macroarea (Nord, Centro e Mezzogiorno) o per l'intero territorio nazionale (Italia), così da ottenere una serie storica dei coefficienti di correlazione dal 2004 al 2020.

Nel complesso, la figura 6 evidenzia un andamento disomogeneo e non lineare, con coefficienti che oscillano in modo significativo da un anno all'altro e tra le diverse aree geografiche, assumendo sia valori positivi che negativi. A livello nazionale, ad esempio, la correlazione è positiva nei primi anni (2004-2006), ma si inverte a partire dal 2007, diventando negativa, con valori che raggiungono circa -0,70. Per quanto riguarda le macroaree geografiche, si osservano dinamiche differenti: nel Nord, la correlazione è debole e prossima allo zero, mentre nel Mezzogiorno emergono correlazioni negative più marcate e persistenti, con punte fino a -0,72. Nel Centro Italia, la correlazione mostra andamenti altrettanto variabili, oscillando

Figura 6. Variazione annuale delle correlazioni tra il Tasso di occupazione senior (TOS) e il Tasso di disoccupazione giovani (TDG) in Italia e per macroaree (Nord, Centro e Mezzogiorno), 2004-2020



Fonte: elaborazioni su dati Istat, 2024.

inizialmente intorno allo zero per poi diventare negativa e consistente, in linea con il Meridione, soprattutto tra il 2012 e il 2018, con un valore minimo di -0,65 nel 2017. Tuttavia, anche in questo caso, i dati variano notevolmente nel corso del tempo.

Alla luce di queste evidenze, appare chiara una relazione non stabile, uniforme o generalizzabile tra TOS e TDG. Il comportamento della variabile TDG potrebbe essere influenzato da fattori esterni non considerati in questa analisi. Ad esempio, la recente crescita del numero di giovani iscritti in università o laureati potrebbe aver ridotto il numeratore del Tasso di disoccupazione giovanile, attenuandone il valore anche in assenza di un effettivo miglioramento del mercato del lavoro. Tale effetto strutturale rende ancora più complessa l'interpretazione della correlazione osservata. Le correlazioni osservate, pur interessanti, non consentono di trarre conclusioni nette. Al contrario, la variabilità dei coefficienti nel tempo e nello spazio suggerisce di considerare l'influenza di molteplici fattori non inclusi in questa analisi, quali la composizione settoriale dell'economia, le politiche attive del lavoro e pensionistiche, e le dinamiche demografiche, educative e migratorie. Di conseguenza, la mera presenza (o il nuovo ingresso) di lavoratori senior, nei diversi contesti territoriali e su scala nazionale, non sembra tradursi in un 'blocco' per l'accesso dei giovani al mercato del lavoro. L'andamento instabile e i valori spesso prossimi allo zero indicano che per comprendere la disoccupazione giovanile è fondamentale considerare un insieme più ampio di fattori, sia strutturali sia congiunturali, piuttosto che limitarsi al solo confronto.

6. Le trasformazioni del mercato del lavoro in Italia

L'analisi delle correlazioni tra TOS e TDG ha permesso di identificare nel 2014 un punto di discontinuità strutturale. Come già evidenziato precedentemente, tale analisi non consente, di per sé, di inferire relazioni causali dirette tra l'occupazione senior e la disoccupazione giovanile.

Per un'interpretazione adeguata della relazione fra le due variabili, risulta pertanto opportuno considerare anche le principali riforme del mercato del lavoro introdotte in Italia nel periodo in esame (2003-2020), che hanno inciso sulle condizioni di accesso, permanenza e uscita dal mercato e, di conseguenza, sugli equilibri intergenerazionali, nonché il più ampio quadro macroeconomico, i cui andamenti – alternando fasi di espansione, contrazione e ristrutturazione – hanno certamente giocato un ruolo determinante.

A partire dalla fine degli anni Novanta, in Italia come in Europa prende avvio un'intensa stagione di riforme di de-regolamentazione del mercato del lavoro, con l'obiettivo di affrontare la disoccupazione, in particolare giovanile e femminile. Gli anni in cui l'Europa adotta politiche di maggiore flessibilità sono segnati, infatti, da un aumento generalizzato dei livelli di disoccupazione. Il dibattito accademico di allora verteva sulla necessità di rimuovere le 'rigidità' del mercato del lavoro – sindacati forti, ampie prestazioni sociali, salari minimi elevati e vincoli ai licenziamenti – considerati di ostacolo alla mobilità dei lavoratori e all'allocazione efficiente delle risorse. In particolare, si sosteneva che l'elevato costo dei licenziamenti scoraggiasse le nuove assunzioni, prolungando i periodi di disoccupazione (Fana *et al*. 2016). Nelle economie periferiche europee si ipotizzava quindi che un incremento della flessibilità interna del mercato del lavoro e una maggiore mobilità degli individui avrebbero ridotto la disoccupazione, migliorato il benessere sociale attraverso una maggiore dinamicità del mercato e, di conseguenza, ristretto il gap competitivo con le economie del centro, inaspritosi a partire dalla crisi del 2008 (Cirillo *et al*. 2016; Fana *et al*. 2016).

Le prime riforme del mercato del lavoro italiano – la legge n. 196/1997, nota come Pacchetto Treu, e la Legge Biagi nel 2003[1] – focalizzate sulla flessibilità del lavoro (Rodano 2015; Fana *et al*. 2016; Lamonica 2018), non sembrano, tuttavia, invertire il trend di disoccupazione giovanile nel primo decennio considerato dalla nostra analisi.

Nel primo decennio del XXI secolo, in Italia si osserva una riduzione del tasso di disoccupazione complessivo, a fronte di un incremento di quello giovanile, che rimane stabilmente al di sopra della media europea e presenta un marcato divario di genere. A seguito della crisi finanziaria del 2008, originatasi negli Stati Uniti, anche il tasso di disoccupazione generale torna a salire, associandosi a fenomeni di (doppia) polarizzazione e deindustrializzazione (Cirillo *et al*. 2016; Lamonica 2018). In questo contesto di crisi economica e di riforme orientate alla flessibilità, l'occupazione senior continua a crescere – sostenuta anche da dinamiche demografiche – mentre la disoccupazione giovanile registra un incremento marcato, contribuendo alla forte correlazione positiva osservata fino al 2014.

In un clima di emergenza, viene varata la legge n. 92/2012, cosiddetta Riforma Fornero, che interviene sulle forme contrattuali flessibili, modifica la disciplina dei licenziamenti, rinnova il sistema di ammortizzatori sociali, rafforza le politiche attive del lavoro e detta norme per promuovere l'occupazione di donne e lavoratori anziani. La riforma, nello specifico, configura il contratto a tempo indeterminato quale contratto prevalente, disincentivando il ricorso ai contratti a tempo determinato. Il tutto, in un contesto dove l'abolizione dell'indennità di disoccupazione e mobilità insieme all'innalzamento dell'età pensionabile producono un 'effetto tappo' all'uscita dal mercato dei lavoratori più anziani, con pesanti conseguenze sull'occupazione giovanile. Sebbene l'analisi statistica punti al 2014 come anno di inversione della correlazione, l'introduzione della Riforma Fornero, che proroga la permanenza dei lavoratori senior nel mercato, rappresenta un momento importante, in quanto, in concomitanza con la crisi del debito sovrano nell'area euro, può aver esacerbato la situazione giovanile e mantenuto la correlazione positiva tra TOS e TDG fino all'emergere di nuovi fattori post 2014.

L'introduzione di una maggiore flessibilità nel mercato del lavoro europeo non mette in secondo piano la necessità di garantire adeguati livelli di sicurezza e sostegno alle fasce più vulnerabili, tra cui disoccupati, giovani in cerca di prima occupazione e NEET (*Not in Education, Employment or Training*), la cui presenza in Europa risultava allora in forte

1 La legge promuove nuove tipologie contrattuali atipiche – il lavoro a chiamata, lo staff leasing e il job sharing – nonché interventi normativi finalizzati a limitare l'utilizzo delle collaborazioni coordinate e continuative (co.co.co.), come alternativa al lavoro subordinato, promuovendone la trasformazione in collaborazioni a progetto. La legge regolamenta altresì forme di lavoro occasionale e accessorio, incentiva l'apprendistato professionalizzante e introduce il contratto d'inserimento, al fine di favorire l'ingresso nel mercato del lavoro dei giovani e delle categorie svantaggiate.

crescita. In questa cornice, trova spazio il modello danese della flexicurity, ampiamente promosso a livello europeo, e capace di coniugare la flessibilità contrattuale con un solido sistema di protezione sociale, assicurando un equilibrio tra dinamismo del mercato del lavoro e tutela dei lavoratori (Orientale Caputo 2021). Al riguardo, il caso italiano si distingue per il carattere emblematico delle sue riforme, che determinano, da un lato, una deregolamentazione selettiva e parziale del mercato del lavoro e, dall'altro, un'accentuazione della polarizzazione e della segmentazione occupazionale (Lamonica 2018). La struttura del mercato del lavoro italiano, quindi, già contraddistinta da elementi di squilibrio e asimmetrie, rimane fortemente frammentata a completamento del processo di liberalizzazione, realizzatosi in maniera significativa nel 2014, con l'introduzione della legge delega n. 183/2014 e successivi decreti attuativi, meglio nota come Jobs Act.

La riforma, promotrice del modello di flexsecurity, introduce i contratti a tempo indeterminato a tutele crescenti per tutti i neoassunti, estende a tutti i lavoratori disoccupati la possibilità di accedere ai sussidi di disoccupazione, incrementandone in modo significativo l'entità, e rafforza gli strumenti finalizzati a facilitare l'incontro tra domanda e offerta di lavoro, potenziando le politiche attive del lavoro. Per facilitare la transizione dei giovani nel mercato del lavoro, il Jobs Act introduce, inoltre, due misure specifiche: la prima, sull'ampliamento dell'apprendistato; la seconda, che rende obbligatoria l'alternanza scuola-lavoro, sul modello del sistema duale tedesco, con l'intento di favorire una maggiore integrazione tra il mondo dell'istruzione e quello lavorativo (Rodano 2015; Cirillo *et al*. 2016; Pinelli *et al*. 2017; Lamonica 2018).

Parallelamente al Jobs Act, viene avviato il Programma Garanzia Giovani – finanziato per 1,5 miliardi di euro, comprensivo di fondi europei e cofinanziamento nazionale – con l'obiettivo di contrastare la disoccupazione giovanile e migliorare l'occupabilità dei giovani NEET attraverso l'offerta di tirocini, corsi di formazione e l'introduzione di incentivi occupazionali (Vesan 2014; Isfol 2016). A fronte delle ambizioni programmatiche, Garanzia Giovani mostra alcuni limiti strutturali. Innanzitutto, la forte pressione sui tempi di spesa europea determina un lancio repentino delle misure, generando un disallineamento tra annunci e implementazione concreta. Il progetto, concepito principalmente 'a tavolino' sul piano della governance, trascura la costruzione di interventi adeguati alle esigenze dei NEET e non prevede sufficienti spazi di adattamento locale; ne deriva un'implementazione frammentaria, avviata solo nella seconda metà del 2015, in concomitanza con l'esaurimento dei finanziamenti. A ciò si aggiunge la carenza di risorse organizzative nei Centri per l'impiego, incapaci di assorbire l'improvviso incremento di utenza. Infine, l'instabilità istituzionale, segnata dalle successive riforme del Jobs Act e dal riassetto delle competenze multilivello, ostacola la coerenza e la continuità necessarie per un apprendimento politico-istituzionale duraturo (Vesan e Lizzi 2016).

Tali interventi, pur con i loro limiti, coincidono temporalmente con il miglioramento del quadro occupazionale – nel periodo successivo al 2014 – e pertanto la loro azione congiunta o parallela potrebbe aver contribuito all'inversione della correlazione osservata. L'analisi di Fana *et al*. (2016), tuttavia, metterebbe in discussione una tale conclusione, in considerazione dell'incapacità del Jobs Act di stimolare in modo significativo l'occupazione complessiva – stabile e qualificata – nonché del forte ricorso al lavoro temporaneo e a strumenti contrattuali atipici, come i voucher. Sebbene si sia registrato un lieve aumento in termini assoluti dei contratti a tempo indeterminato, questo non ha compensato la crescita dei contratti a termine, che, in termini relativi, hanno avuto un incremento significativamente maggiore. Inoltre, l'aumento dei contratti a tempo indeterminato è derivato principalmente dalla trasformazione di contratti preesistenti in contratti più stabili, piuttosto che dalla creazione di nuove posizioni lavorative. Infine, l'aumento dei contratti stabili ha interessato prevalentemente le coorti di età superiore ai 55 anni, a scapito dei giovani. La lieve flessione del tasso di disoccupazione nel 2015 è da ricondurre, quindi, più all'ingresso degli inattivi nel mercato del lavoro – non considerati nella nostra analisi – grazie, ad esempio, a Garanzia Giovani, che agli effetti diretti delle riforme. Rodano (2015), di contro, sottolinea come l'obiettivo principale del Jobs Act non fosse tanto la creazione netta di nuovi posti di lavoro, quanto la riprofilazione della composizione contrattuale a favore dei rapporti a tempo indeterminato con tutele crescenti. Ad ogni modo, la limitata efficacia del Jobs Act

nel promuovere stabilità e nuova occupazione evidenzia la complessità del quadro e suggerisce come l'inversione della correlazione nel secondo periodo sia dovuta a una combinazione di fattori che includono anche una ripresa economica, seppur modesta, della zona euro (World Bank 2015).

In ultima analisi, per comprendere appieno l'andamento occupazionale della fascia dei lavoratori senior e il suo rapporto con la disoccupazione giovanile, è essenziale considerare il decreto-legge n. 4/2019, noto come Quota 100, introdotto dal Governo Conte I con la Legge di Bilancio 2019. La misura, realizzata con l'obiettivo dichiarato di ridurre il tasso di disoccupazione dei giovani, intende favorirne l'ingresso nel mercato del lavoro attraverso il pensionamento anticipato dei lavoratori più anziani.

In sintesi, il rafforzamento delle politiche attive del lavoro, l'introduzione dei contratti a tutele crescenti e misure di transizione scuola-lavoro, seppur con evidenti limiti, possono aver migliorato le condizioni di accesso al mercato del lavoro per i giovani nel periodo successivo al 2014, mentre le politiche previdenziali – quali la Riforma Fornero e Quota 100 – hanno inciso rispettivamente sulla permanenza o sull'uscita dei lavoratori anziani. Questi interventi normativi e le dinamiche economiche sottostanti possono spiegare il cambiamento nella correlazione tra TOS e TDG osservato nel 2014, modificando la relazione da positiva a negativa. La relazione non si presenta stabile nel tempo né uniforme tra le macroaree geografiche, suggerendo che fattori locali, settoriali e istituzionali continuano a esercitare un ruolo cruciale nel mercato del lavoro italiano e concorrono nel modulare la relazione tra occupazione senior e disoccupazione giovanile nei diversi contesti.

7. Implicazioni e conclusioni

La relazione tra il TOS e il TDG evidenzia dinamiche complesse che sfidano la narrazione di un conflitto generazionale. Non vi sono evidenze empiriche che dimostrino una correlazione diretta e stabile tra l'occupazione dei lavoratori senior e la disoccupazione giovanile, malgrado la crescita dell'occupazione tra i primi e il prolungamento della loro permanenza nel mercato del lavoro. Al contrario, è più plausibile attribuire le difficoltà occupazionali dei giovani a politiche del lavoro inefficaci e a criticità strutturali del mercato del lavoro italiano.

Com'è noto, l'ingresso dei giovani nel mercato del lavoro è condizionato tanto dall'impianto normativo, che ne regola e supporta il processo di transizione verso l'occupazione, quanto dal grado di integrazione tra il sistema formativo e il tessuto produttivo (Lamonica 2018). A partire dal 2014 – anno chiave nella nostra analisi – il mercato del lavoro italiano conosce timidi cambiamenti determinati da fattori economici (Istat 2015; 2016), normativi e sociali, che lasciano, tuttavia, inalterata la sua struttura. L'uscita dalla fase più critica della crisi economica rappresenta un punto di svolta che favorisce una lenta ripresa dell'occupazione. In questo contesto, riforme del mercato del lavoro, quali il Jobs Act, concepite per flessibilizzare il lavoro, garantendo tuttavia sicurezza e stabilità occupazionale, mostrano limiti importanti. Sebbene si registri un aumento dei contratti a tempo indeterminato, gran parte di essi riguarda la stabilizzazione dei rapporti preesistenti piuttosto che la creazione di nuova occupazione. Inoltre, il mercato del lavoro continua a penalizzare i giovani e le categorie più vulnerabili, con una crescita occupazionale concentrata principalmente nelle coorti più anziane. Anche il programma Garanzia Giovani, progettato per contrastare la disoccupazione giovanile e migliorare l'occupabilità dei giovani NEET, si rivela inefficace a livello sistemico. Inoltre, la perdurante disparità occupazionale tra le macroaree geografiche considerate (Nord, Centro, Mezzogiorno) evidenzia come le politiche adottate non abbiano adeguatamente considerato le profonde disuguaglianze territoriali, compromettendone l'efficacia complessiva.

Nonostante i dati indichino l'assenza di un vero e proprio conflitto intergenerazionale, potrebbero permanere percezioni distorte di competizione tra gruppi anagrafici, specialmente in un mercato del lavoro strutturalmente sbilanciato verso le fasce d'età più avanzate e caratterizzato, come già reso noto in precedenza, da marcate disuguaglianze occupazionali tra territori (Reyneri 2017; Orientale Caputo 2021).

Pur restando necessarie politiche mirate sia a sostenere l'occupazione giovanile sia a ridurre le disparità territoriali, la comunità scientifica e le istituzioni pubbliche hanno già cominciato a lavorare

in modo strutturato sulle trasformazioni demografiche e sulle sfide poste dall'invecchiamento attivo. Il programma Age-It[2], promosso nell'ambito del Piano nazionale di ripresa e resilienza (PNRR), Missione 4 – Istruzione e Ricerca, Componente 2, ad esempio, intende trasformare l'Italia in un polo scientifico internazionale per lo studio dell'invecchiamento. Nel dettaglio, il progetto, che beneficia di un ingente finanziamento e coinvolge una vasta rete di università, centri di ricerca, imprese, istituzioni pubbliche e attori della società civile, adotta un approccio fortemente interdisciplinare, che integra competenze in ambiti quali demografia, geriatria, *data science*, sociologia, diritto, economia e altre discipline. Con l'obiettivo di promuovere un invecchiamento inclusivo e sostenibile, il progetto è organizzato attorno a tre assi di ricerca e dieci *spoke* tematici, orientati allo sviluppo di soluzioni socioeconomiche, biomediche e tecnologiche (Boffo 2022). Parallelamente, sul piano europeo, il programma COST Action LeverAge – promosso dalla Commissione europea e sostenuto da oltre 150 ricercatori – investe in modo sinergico in queste stesse tematiche, riconoscendo come i rapidi mutamenti demografici richiedano risposte innovative, integrate e coordinate (Marcus *et al*. 2024; Low e Pok 2025).

Inoltre, sul piano delle azioni concrete da realizzare per mitigare la percezione di una presunta concorrenza tra senior e giovani e favorire un reciproco scambio, risulta opportuno investire in pratiche di Diversity Management, ossia di gestione della diversità generazionale. Quest'ultimo, infatti, non solo mira a promuove il benessere organizzativo, ma rappresenta altresì un fattore chiave per il successo, produttivo e organizzativo, aziendale (Riccio e Scassellati 2008; Ali e French 2019; Romano 2020; Basaglia *et al*. 2022; Marcus *et al*. 2024). In questo senso, il mentoring intergenerazionale e politiche *Human Resources* (HR) inclusive, uniti alla valorizzazione del capitale umano, possono promuovere il trasferimento di competenze, garantire aggiornamento professionale e benessere lavorativo, riconoscendo il contributo unico di ogni generazione e favorendo benessere, innovazione e senso di appartenenza (Ali e French 2019; Froidevaux *et al*. 2020; Boehm *et al*. 2021; Firzly *et al*. 2021; Fan *et al*. 2023; Wang *et al*. 2024).

Per quanto riguarda le analisi delle correlazioni condotte, occorre tenere conto di alcune criticità metodologiche che possono limitare la portata delle evidenze empiriche. Uno dei principali vincoli del presente studio riguarda l'utilizzo dei dati Istat, la cui definizione di 'occupato' include chiunque abbia svolto anche una sola ora di lavoro retribuito nel periodo di riferimento. Tale criterio, per quanto diffuso nelle rilevazioni ufficiali, non consente di cogliere con precisione le differenze tra forme di impiego stabile o precario, né di valutare la reale qualità dell'occupazione registrata. Questa lacuna informativa complica ulteriormente l'interpretazione dell'apparente assenza di un conflitto tra generazioni, poiché non permette di approfondire le condizioni contrattuali e reddituali che distinguono le diverse coorti occupate.

Alla luce di questi limiti, futuri approfondimenti dovrebbero concentrarsi su un'analisi più puntuale della natura e durata dei contratti, sulla frequenza delle occupazioni discontinue e sulle traiettorie della popolazione inattiva nel tempo. Un approccio di questo tipo consentirebbe non solo di valutare in modo più esaustivo l'efficacia delle riforme del lavoro, ma anche di comprendere con maggiore accuratezza come i cambiamenti normativi ed economici abbiano influenzato le reali opportunità occupazionali e la qualità del lavoro disponibile per le diverse generazioni. Inoltre, risulterebbe opportuno sviluppare e analizzare più sistematicamente le pratiche di Diversity Management, in particolare quelle orientate alla gestione della diversità generazionale, come leva strategica per affrontare a livello di impresa le sfide di un mercato del lavoro in rapido invecchiamento.

Bibliografia

Ali M., French E. (2019), Age diversity management and organisational outcomes: The role of diversity perspectives, *Human Resource Management Journal*, 29, n.2, pp.287-307

Basaglia S., Cuomo S., Simonella Z. (2022), *L'organizzazione inclusiva: pari opportunità e diversity management*, Milano, Egea

2 *Ageing well in an ageing society. A novel public-private alliance to generate socioeconomic, biomedical and technological solutions for an inclusive Italian ageing society (2023-2026)*. Per maggiori informazioni: https://ageit.eu/wp/ (consultato il 20 aprile 2025).

Behaghel L., Caroli E., Roger M. (2014), Age-biased technical and organizational change, training and employment prospects of older workers, *Economica*, 81, n.322, pp.368-389

Belloni M., Maccheroni C. (2013), Actuarial fairness when longevity increases: An evaluation of the Italian pension system, *The Geneva Papers on Risk and Insurance. Issues and Practice*, 38, n.4, pp.638-674

Benassi F., Busetta A., Gallo G., Stranges M. (2021), Diseguaglianze tra territori, in Billari F.C., Tomassini C. (a cura di), *Rapporto sulla popolazione. L'Italia e le sfide della demografia*, Bologna, il Mulino, pp.135-162

Billari F.C. (2023), *Domani è oggi. Costruire il futuro con le lenti della demografia*, Milano, EGEA

Billari F.C., Tomassini C. (2024), *Rapporto sulla popolazione: l'Italia e le sfide della demografia*, Bologna, il Mulino

Boehm S.A., Schröder H., Bal M. (2021), Age-related human resource management policies and practices: Antecedents, outcomes, and conceptualizations, *Work, Aging and Retirement*, 7, n.4, pp.257-272

Boffo V. (2022), Active ageing: Il ruolo dell'apprendimento permanente. Editoriale, *Epale Journal on Adult Learning and Continuing Education*, n.12, pp.1-85

Burke R.J., Cooper C.L., Field J. (2013), The aging workforce: Individual, organizational and societal opportunities and challenges, in Field J., Burke R. J., Cooper C.L. (eds.), *The SAGE Handbook of Aging, Work and Society*, London, SAGE Publications, pp.1-20

Cazzola G. (2004), *Lavoro e welfare: giovani versus anziani: conflitto tra generazioni o lotta di classe del XXI secolo?,* Soveria Mannelli, Rubbettino Editore

Charness G., Villeval M.C. (2009), Cooperation and competition in intergenerational experiments in the field and the laboratory, *American Economic Review*, 99, n.3, pp.956-978

Cirillo V., Fana M., Guarascio D. (2016), Did Italy need more labour flexibility?, *Intereconomics: Review of European Economic Policy*, 51, n.2, pp.79-86

Della Porta D., Keating M., Pianta M. (2021), Inequalities, territorial politics, nationalism, *Territory, Politics, Governance*, 9, n.3, pp.325-330

Fan P., Song Y., Fang M., Chen X. (2023), Creating an age-inclusive workplace: The impact of HR practices on employee work engagement, *Journal of Management & Organization*, 29, n.6, pp.1179-1197

Fana M., Guarascio D., Cirillo V. (2016), La crisi e le riforme del mercato del lavoro in Italia: Un'analisi regionale del Jobs Act, *Argomenti*, 3, n.5, pp.5-30

Firzly N., Van de Beeck L., Lagacé M. (2021), Let's work together: Assessing the impact of intergenerational dynamics on young workers'ageism awareness and job satisfaction, *Canadian Journal on Aging/La Revue Canadienne du Vieillissement*, 40, n.3, pp.489-499

Froidevaux A., Alterman V., Wang M. (2020), Leveraging aging workforce and age diversity to achieve organizational goals: A human resource management perspective, in Czaja S.J., James J., Grosch J., Sharit J. (eds.), *Current and emerging trends in aging and work*, Cham, Springer, pp.33-58

Garcìa-Pereiro T., Pace R., Venere G. (2024), Giovani impegnati nella transizione verso lo stato adulto: dati e confronti, in Conte M., Rubino R. (a cura di), *Nuovi orizzonti di ricerca nella società che cambia. Popolazione, formazione e tecnologie*, Bari, Cacucci Editore, pp.205-236

Inapp, Aversa M. L., Checcucci P., Iadevaia V. (a cura di) (2025), *Digitalizzazione e invecchiamento della forza lavoro nelle piccole e medie imprese italiane,* Inapp Report n.56, Roma, Inapp

Isfol (2016), *Rapporto sulla Garanzia Giovani in Italia*, Roma, Isfol

Istat (2024), *Rapporto Annuale 2024. La situazione del Paese*, Roma, Istat

Istat (2019), *Rapporto Annuale 2019. La situazione del Paese*, Roma, Istat

Istat (2016), *Rapporto Annuale 2016. La situazione del Paese*, Roma, Istat

Istat (2015), *Rapporto Annuale 2015. La situazione del Paese*, Roma, Istat

Lamonica V. (2018), Giovani e mercato del lavoro: un'analisi critica della letteratura, *Quaderni IRCrES*, 5, n.31, pp.1-24

Low M.P., Pok W.F. (2025), Global aging workforce: A micro-meso-macro approach, *Activities, Adaptation & Aging*, 49, n.1, pp.1-30

Maccheroni, C., Mignolli, N., Pace, R., & Venere, G. (2025). Older Adult Surge and Social Welfare Inequalities in Italy: The Impact of Population Ageing on Pensions and the Welfare System. *Populations*, 1(2), 9.

Marcaletti F. (2013), La dinamica intergenerazionale nei mercati del lavoro: Tra conflitto, mutua esclusione e misure per l'inclusività, *Studi di Sociologia,* 51, n.3/4, pp.307-316

Marcus J., Scheibe S., Kooij D., Truxillo D. M., Zaniboni S., Abuladze L., Al Mursi N., Bamberger P. A., Balytska M., Betanzos N. D., Perek-Białas J., Boehm S. A., Burmeister A., Cabib I., Caon M., Deller J., Derous E., Drury L., Eppler-Hattab R., Fasbender U., Fülöp M., Furunes T., Gerpott F. H., Goštautaitė B., Halvorsen C. J., Hernaus T., Inceoglu I., Iskifoglu M., Ivanoska K. S., Kanfer R., Kenig N., Kiran S., Klimek S., Kunze F., Mertan E. B., Varianou-Mikellidou C., Moasa H., Ng Y. L., Parker S. K., Reh S., Resuli V., Schmeink M., Silberg S., Sousa I. C., Steiner D. D., Stukalina Y., Tomas J., Topa G., Turek K., Vignoli M., von Bonsdorff M., Wang D., Wang M., Yeung D. Y., Yildirim K., Zhang X., e Žnidaršič J. (2024), LeverAge: A European network to leverage the multi-age workforce, *Work, Aging and Retirement,* 10, n.4, pp.309-316

Orientale Caputo G. (2021), *Analisi sociale del mercato del lavoro*, Bologna, il Mulino

Pinelli D., Torre R., Pace L., Cassio L., Arpaia A. (2017), *The recent reform of the labour market in Italy: A review,* Discussion Paper n.072, Luxembourg, Publications Office of the European Union

Reyneri E. (2017), *Introduzione alla sociologia del mercato del lavoro*, Bologna, il Mulino

Riccio G., Scassellati A. (2008), *Le politiche aziendali per l'age management. Materiali per un piano nazionale per l'invecchiamento attivo*, Monografie sul mercato del lavoro e le politiche per l'impiego n.1/2008, Roma, Isfol

Ripamonti S., Bruno A., Galuppo L., Kaneklin C. (2021), Le forme di scambio tra le generazioni nei contesti organizzativi: La transizione dei neolaureati al mondo del lavoro, *Ricerche di psicologia*, 3, pp.1-33

Rodano G. (2015), Il mercato del lavoro italiano prima e dopo il Jobs Act, *www.bollettinoadapt.it,* 8 maggio

Romano A. (2020), *Diversity & disability management: Esperienze di inclusione sociale*, Milano, Mondadori

Rosina A. (a cura di) (2024), *Demografia e forza lavoro in Italia: Rapporto a cura del consigliere Alessandro Rosina*, Roma, Cnel

Rudolph C.W., Zacher H. (2015), Intergenerational perceptions and conflicts in multi-age and multigenerational work environments, in Finkelstein L.M., Truxillo D.M., Fraccaroli F., Kanfer R. (eds.), *Facing the challenges of a multi-age workforce: A use-inspired approach*, New York, Routledge/Taylor & Francis Group, pp.253-282

Socci M. (2015), Giovani e anziani nel mercato del lavoro tra solidarietà e conflitto, *Prisma: economia, società, lavoro*, n.3, pp.54-71

Struffolino E., Filandri M. (2021), Povertà e ricchezza tra le famiglie di lavoratori in Italia: Trent'anni di svantaggio cumulativo, *Sociologia del lavoro*, 161, n.3, pp.97-121

Vesan P. (2014), La Garanzia Giovani: Una seconda chance per le politiche attive del lavoro in Italia?, *Politiche Sociali*, n.3, pp.377-396

Vesan P., Lizzi R. (2016), La Garanzia Giovani in Italia e l'approccio del new policy design: Tra aspettative, speranze e delusioni, *Rivista Italiana di Politiche Pubbliche*, n.1, pp.5-38

Viesti G. (2021), *Centri e periferie: Europa, Italia, Mezzogiorno dal XX al XXI secolo*, Bari, Laterza

Vignoli D., Tomassini C. (2023), *Rapporto sulla popolazione: Le famiglie in Italia: Forme, ostacoli, sfide*, Bologna, il Mulino

Wang Z., Fu J., Bai W. (2024), Public management approaches to an aging workforce: Organizational strategies for adaptability and efficiency, *Frontiers in Psychology*, n.15, 1439271

World Bank (2015), *EU regular economic report: Modest recovery, global risk*. *Document of the World Bank*, Washington DC, World Bank Group

Giuseppe Venere
giuseppe.venere@uniba.it
Dottorando di ricerca in Scienze politiche e sociali per la sicurezza e lo sviluppo presso l'Università degli Studi di Bari Aldo Moro, con principale ambito di interesse l'invecchiamento della popolazione (STAT-03/A). È un assistente sociale specialista e valutatore ospite presso l'Agenzia regionale strategica per la salute e il sociale della Puglia per un progetto regionale sull'invecchiamento attivo. Fra le pubblicazioni recenti si segnala: Older adult surge and social welfare inequalities in Italy: the impact of population ageing on pensions and the welfare system, *Populations MDPI*, 2025.

Giorgia Panico
giorgia.panico@unisalento.it
Dottoranda di ricerca in Scienze umane e sociali (GSPS-08/A), curriculum Scienze sociali e politiche, presso l'Università del Salento. È Cultore della materia in sociologia economica e del lavoro presso l'Università degli Studi di Bari Aldo Moro. Fra le pubblicazioni recenti si segnala: Salute in Umbria. Come livelli diversi di governo affrontano rischi e minacce nella municipalità di Terni: il Documento Unico di Programmazione, U3 – *UrbanisticaTre*, 2025.

Scaffale - *Rubrica di recensioni*

Geopolitica dell'Intelligenza artificiale

Alessandro Aresu

Milano, Feltrinelli, 2024, pp.576

Finalista al Premio Strega Saggistica 2025, il libro di Aresu si articola in numerosi capitoli imbevuti di storytelling, fenomeno peraltro assai in voga nell'epoca attuale, a partire dai titoletti assai accattivanti – *Dal Kentucky a Lady Gaga; Morris, il mio eroe; Cynar*; solo per citarne alcuni – in cui vengono sostanzialmente delineati i principali protagonisti dell'attuale rivoluzione tecnica dell'IA. Non solo di quella generativa (GenAI), ma soprattutto di quella che viene definita come generale (AGI), a partire dalla volgarizzazione fattane, intorno al 2002, da Shane Legg.

Uno degli aspetti più interessanti e originali dell'opera di Aresu è proprio quest'approccio 'biografico-critico' alla storia recente dell'IA. Il libro non si limita, difatti, a una trattazione solo tecnica oppure geopolitica: al contrario, intreccia le traiettorie personali e professionali dei maestri riconosciuti, tuttora viventi, delle moderne *corporation* tecnologiche le quali, come tutte le antiche corporazioni, mantengono gelosamente custoditi non solo i propri segreti algoritmici, ma anche i propri 'fossati' competitivi (i *wide moat* di Warren Buffett): dalle figure più note di Morris Chang, Elon Musk, Peter Thiel, passando per quelle meno conosciute dal grande pubblico quali Ian Buck, Bryan Catanzaro, William 'Billy' Dally, e molti altri. Questi veri e propri demiurghi moderni hanno fondato big tech del calibro di DeepMind, OpenAI, NVIDIA, Taiwan Semiconductor Manufacturing Company (TSMC), Mellanox, ecc., oppure hanno rivitalizzato vecchi colossi, quali Microsoft, che sembravano inesorabilmente avviati verso una sempre maggiore insignificanza tecnologica.

Questa scelta si configura come uno dei punti di forza dell'opera, offrendo al lettore strumenti critici per comprendere l'IA non come entità astratta o mitologica, bensì come campo vivo tuttora attraversato da tensioni, passioni, rivalità (notoria quella tra Altman e Musk), divergenze etiche e biografie talvolta segnate da fragilità, ma sempre da operosità e tenacia. Aresu mostra che innovazione e leadership emergono non solo da competenze tecnico-scientifiche e vision imprenditoriale, ma anche da reti di collaborazione – così diffuse nella Silicon Valley – fallimenti e intuizioni futuristiche nell'immaginare nuovi prodotti scalabili e, di conseguenza, letteralmente creare *ex nihilo* mercati divenuti presto globali. Costruttori di un ecosistema digitale che l'Autore definisce come una nuova "fusione militare-tecnologica" riecheggiando il complesso politico-militare-industriale di *eisenhoweriana* e *wrightmillsesiana* memoria. L'autore fa uso di una prosa attenta, affabulatoria, mai banale, da erudito di altri tempi, con continue citazioni filosofiche, letterarie e sociologiche – a partire da Massimo Cacciari a cui è dedicato il libro – con note bibliografiche che occupano ben 91 pagine finali. Tuttavia, riesce nel fine che si era proposto, ossia quello di mettere in luce l'umanità e le contraddizioni di chi ha plasmato il destino tecnologico degli ultimi decenni, riuscendo a restituire ai padri dell'IA una dimensione umana che va oltre l'icona mediatica o la narrazione agiografica.

Su questa trama narrativa si innesta il vero tema del libro, vale a dire la non mai sopita tensione geopolitica, in un'ottica di sicurezza nazionale, che oggi fa registrare un vero e proprio parossismo con il capitalismo politico dell'Amministrazione Trump. L'IA, "l'invenzione definitiva dell'umanità", questo l'incipit del libro, è divenuta, difatti, una leva strategica fondamentale nella ridefinizione degli equilibri mondiali. La competizione tra Stati Uniti e Repubblica Popolare Cinese rappresenta il fulcro della silente guerra tecnologica, e oggigiorno, di quella commerciale basata sui dazi: entrambe le potenze vedono nell'IA non solo un motore di crescita economica, ma anche uno strumento di dominio geopolitico e militare. Gli USA mantengono il primato negli investimenti privati e nell'innovazione grazie a big tech quali OpenAI, Google, Microsoft e Nvidia, mentre la Cina punta su una strategia governativa centralizzata, e assai raffinata, imperniata sia sull'ulteriore sviluppo della potenza manifatturiera sia sullo strategico "vantaggio dell'arretratezza" – concetto teorico

sviluppato da Alexander Gerschenkron – esemplificata da colossi quali Huawei, BYD, Baidu, Alibaba e Tencent, per nominare solo i più conosciuti, con l'obiettivo dichiarato di raggiungere la leadership globale entro il 2030. Questa corsa all'IA coinvolge anche l'Unione europea, il Giappone, la Corea del Sud e i Paesi del Golfo, ognuno con specificità e limiti. L'Ue, in particolare, cerca di affermarsi come regolatore globale, promuovendo standard etici e di sicurezza (*AI Act*, *AI Continent Action Plan*), ma fatica a competere in termini di scala industriale e innovazione. L'Autore dedica solo qualche breve cenno derisorio alla postura comunitaria in tema di superfetazione regolatoria: "Siamo morti che camminano. Per darci un tono, ci definiamo sonnambuli" (p. 439). Oppure, in maniera ancora più caustica: "Diventiamo il paese del futuro con la ristrutturazione delle facciate degli edifici? O abbiamo altri progetti?" (p. 441). Il libro di Aresu cerca, dunque, di riflettere criticamente su questo panorama in rapido mutamento, dove la corsa globale all'IA non è solo una gara tecnologica, ma una sfida sistemica che ridefinisce potere, economia e società tra le varie potenze globali.

Un terzo concetto cardine emerso dall'analisi del testo riguarda la natura profondamente pervasiva dell'informazione digitale nella società contemporanea: l'IA è ormai divenuta la 'tessitura invisibile' della vita quotidiana, influenzando la comunicazione, il lavoro, le relazioni e persino le identità personali e collettive. L'Autore spiega come la capillarità degli algoritmi, la crescita esponenziale dei dati e la presenza invisibile delle infrastrutture cloud e di calcolo abbiano reso l'interazione con l'informazione digitale un'esperienza onnipresente e quasi inevitabile. La dematerializzazione di attività, spazi e persino rapporti sociali emerge nel libro come cartina di tornasole delle nuove forme di potere e delle vulnerabilità emergenti: la quotidianità è filtrata dall'architettura invisibile di codici, server, reti e regolamenti, il cui controllo è al centro degli equilibri geopolitici. Pur riconoscendo rischi concreti – alienazione, manipolazione, crisi dei modelli di lavoro tradizionali e disuguaglianze globali – l'Autore evidenzia che tali rischi non condizionano le traiettorie del progresso tecnologico, le quali restano dominate dalla velocità esponenziale dell'innovazione, esemplificata dalla legge di Moore e dalle successive teorizzazioni di Kurzweil sui "ritorni accelerati" e di Jensen Huang sulle GPU (*Graphics Processing Unit)*. L'IA, quindi, rappresenta un rischio reale ma fatalmente ininfluente sul corso complessivo dello sviluppo tecnologico. In definitiva, l'approccio scelto da Aresu è stato quello di aver approfondito le tre tendenze strutturali dell'ora presente: "lo spostamento della manifattura verso l'Asia orientale, la digitalizzazione dell'economia e della società, la pressione politica attraverso la sicurezza nazionale" (p. 438). L'opera fornisce diversi strumenti critici per interpretare questa realtà laddove tecnologia, informazione e vita quotidiana sono ormai inscindibili tratteggiando, così, non solo una storia della tecnologia e delle dinamiche geopolitiche dell'IA, ma anche una riflessione sulla condizione umana nell'era degli algoritmi. Il libro, infine, è arricchito da numerosi casi concreti, cronologie e approfondimenti che aiutano il lettore a districarsi nella cospicua complessità della materia. In ultimo, una nota critica può essere individuata nell'ampiezza del cast biografico chiamato a recitare la parte del protagonista in questo testo: la densità della narrazione rischia talvolta di distrarre dalla linea analitica centrale, generando un racconto stratificato dove il lettore meno esperto potrebbe rischiare di perdersi e non arrivare mai alla fine, trattandosi, pur sempre, di ben 563 pagine complessive. Da questo punto di vista, non vi sono alternative plausibili: o si rimane avvinti, quasi irretiti, com'è successo allo scrivente, dai molteplici fili che l'Autore riesce a tessere con maestria oppure si rischia di far annaspare il lettore in un universo letterario, filosofico, economico, giuridico e sociologico vasto tanto quanto la Wikipedia attuale. A quest'ultimo, tuttavia, si può sempre consigliare l'ausilio dell'IA generativa: ChatGPT, Copilot, Grok, Claude oppure Gemini, poco importa!

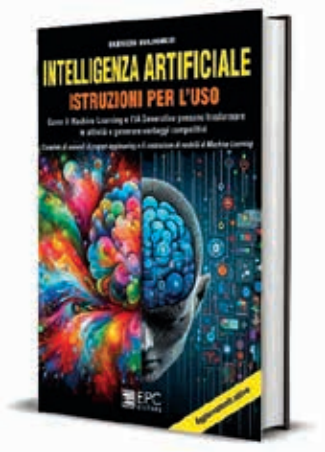
Achille Pierre Paliotta
INAPP

Intelligenza artificiale. Istruzioni per l'uso

Come il Machine Learning e l'IA Generativa possono trasformare le attività e generare vantaggi competitivi

Fabrizio Bolognesi

Roma EPC Editore, 2024, pp.240



Il libro di Fabrizio Bolognesi è senz'altro un testo divulgativo che offre una panoramica accessibile sull'evoluzione dell'intelligenza artificiale (IA) dagli anni Cinquanta a oggi, con un focus particolare su Machine Learning (ML) e IA generativa. L'Autore, infatti, spiega in modo intuitivo questi concetti complessi iniziando con l'analisi di applicazioni usate quotidianamente come Airbnb, Netflix, Google Maps e Amazon, passando poi ad efficaci esempi da mettere in partica durante o dopo la lettura del libro. Pur partendo da un background professionale altamente qualificato[1], l'Autore adotta un linguaggio fluido e accessibile anche su temi complessi, accompagnato da numerose illustrazioni che rendono la lettura agevole anche per chi si avvicina per la prima volta al mondo dell'IA.

Il libro si sviluppa in due parti distinte: *Capire e usare l'IA* (I) e *Manuale di prompt-engineering* (II).

Il capitolo introduttivo della prima parte ripercorre la storia dell'IA, dal Test di Turing[2] degli anni Cinquanta del secolo scorso fino alla sua 'democratizzazione' avvenuta con ChatGPT che a novembre 2022 ha portato l'IA generativa sulle scrivanie di tutti. Seguono due capitoli fondamentali dedicati al ML e all'IA generativa. Un elemento distintivo della prima parte del libro è l'introduzione del 'Binary Bistrot', un'azienda immaginaria utilizzata come laboratorio per applicare i concetti discussi e mostrare attraverso diversi casi d'uso come l'IA possa innovare i processi aziendali, prevedere gli incassi, predire la fedeltà dei clienti e persino creare strategie per aumentare il fatturato. Gli esempi pratici del caso di studio Binary Bistrot semplificano la comprensione di concetti complessi come l'apprendimento supervisionato e non supervisionato e le tecniche di regressione, classificazione e clusterizzazione.

Nel primo esempio il lettore viene accompagnato passo passo nella scrittura di un semplice algoritmo di ML in linguaggio Python[3] in grado di prevedere il fatturato del Binary Bistrot. Per la scrittura dell'algoritmo non è necessario conoscere la programmazione: nell'esempio viene usata la piattaforma Google Collab che permette l'esecuzione di script Python direttamente nel browser dell'utente, rendendo questa esercitazione di facile comprensione. Seguono poi alcuni esempi in cui l'autore introduce le piattaforme 'No Code AI', piattaforme gratuite di ML che non richiedono la scrittura del codice. Il terzo capitolo è interamente dedicato all'IA generativa ed è ricco di suggerimenti su come interagire efficacemente con ChatGPT.

L'ultimo capitolo della prima parte del libro – *Implicazioni dell'IA nella società e nel mondo del lavoro* – affronta importanti questioni etiche legate all'impatto dell'intelligenza artificiale in ambito sociale e lavorativo, con particolare riguardo al rispetto dei diritti individuali. Partendo dalla *Rome Call for AI Ethics*[4] (2021) che esamina la contrapposizione tra l'algocrazia e l'algoretica[5], passando poi per l'*AI Act* dell'Unione

1 Laureato in ingegneria aerospaziale e titolare di un MBA in gestione d'impresa, Bolognesi vanta anche una formazione specifica sull'intelligenza artificiale conseguita presso il Massachusetts Institute of Technology (MIT) di Boston. Vanta un'esperienza ventennale in aziende multinazionali con ruoli di responsabilità nelle aree di sviluppo business, con un solido background in Data Science e competenze nell'applicazione dell'IA in ambito aziendale.

2 Alan Turing (1912-1954) è stato un matematico, crittografo e teorico della computazione britannico. Famoso al grande pubblico per aver decriptato il codice di comunicazione nazista 'Enigma' durante la II guerra mondiale, è considerato uno dei padri fondatori dell'IA. Il celebre 'Test di Turing' rimane ad oggi uno dei metodi più noti per definire e capire l'IA.

3 Python è attualmente il linguaggio di programmazione più usato nei campi di Data Science e Machine Learning.

4 Si veda https://www.romecall.org/.

5 L'algoretica è il campo di studio che si interroga su come gli algoritmi possano essere governati da norme etiche.

europea (2024), l'Autore giunge alla conclusione che la società contemporanea si trovi di fronte a un nuovo fenomeno di divario digitale che "riflette la disparità nelle capacità di comprendere, accedere e utilizzare efficacemente i sistemi di IA, divenuti sempre più centrali in molti aspetti della vita quotidiana" (p. 152).
La parte seconda del libro è un manuale di prompt-engineering per interagire efficacemente con i Large Language Models (LLM) come ChatGPT. Fornisce indicazioni per costruire modelli di IA senza necessità di programmazione. Il paragrafo introduttivo offre una definizione del termine prompt-engineering come "[…] l'arte di formulare domande in modo efficiente per ottimizzare le risposte in funzione delle nostre esigenze." (p. 161). Nei paragrafi successivi troviamo le regole di base del prompt-engineering e le istruzioni per formulare meglio le nostre richieste al modello dell'AI generativa[6] con lo scopo di massimizzare risultati: per l'Autore, nelle interazioni con un LLM, sapere come chiedere è importante quanto sapere cosa chiedere. Sostiene che la capacità dell'utente di strutturare prompt efficaci sia una competenza fondamentale e suggerisce di prestare particolare attenzione agli elementi del prompt che possano indirizzare il modello verso una risposta più accurata e pertinente. Tra diversi elementi possibili da inserire nel prompt vengono individuati come più importanti il ruolo, il compito e il formato. Infatti, a una stessa domanda potremmo ottenere dal modello risposte molto differenti in funzione del ruolo da svolgere (ad esempio, marketing manager, assistente di ricerca o analista finanziario), compito assegnato (ad esempio, riassumere, generare o analizzare un testo) e il formato desiderato dell'output (ad esempio, abstract, report o articolo per un sito web). Nei capitoli a seguire troviamo una rassegna di applicazioni pratiche di prompt-engineering nella generazione, revisione e traduzione di testi e la trasformazione delle idee testuali in immagini e storyboard per video.
In un'intervista, Bolognesi sottolinea che il libro è destinato a un pubblico ampio, dai manager aziendali ai neofiti, con l'obiettivo di spiegare concetti complessi in modo semplice e pratico. In sintesi, il testo rappresenta una guida pratica per comprendere e applicare l'IA in vari contesti, offrendo strumenti utili per sfruttarne le potenzialità nel quotidiano e nel mondo del lavoro. *Intelligenza artificiale. Istruzioni per l'uso* è, in definitiva, un testo che mantiene le promesse, un libro per tutti: accessibile ai non addetti ai lavori per compiere primi passi verso l'uso consapevole del ML e l'IA generativa e, in maniera uguale, utilizzabile da chi ha le basi solide in ICT, statistica, Data Science o Business Intelligence. Dopo la lettura del testo, il lettore sarà in grado di scrivere ed eseguire semplici algoritmi di ML in Python, usare consapevolmente le piattaforme 'No Code AI' e interagire efficacemente con i principali LLM.
Vista la rapidità senza precedenti con cui l'IA sta facendo progressi, influenzando al contempo sia il tessuto sociale che quello lavorativo, risulta particolarmente interessante e utile la possibilità di scaricare aggiornamenti e approfondimenti del libro elaborati dopo la chiusura del testo, disponibili gratuitamente sul sito web dell'EPC Editore.

Boris Sofronic
INAPP

6 La maggioranza degli esempi esposti nel libro è basata su ChatGPT della OpenAI, ad oggi il modello più diffuso dell'IA generativa. Tuttavia, l'Autore non perde mai di vista le alternative come Claude, Copilot, Llama, Gemini e altri LLM.

Decidere con l'IA. Intelligenze artificiali e naturali nel diritto

Amedeo Santosuosso, Giovanni Sartor

Bologna, il Mulino, 2024, pp.226

Il volume si inserisce nell'ormai ampio filone tematico dedicato all'analisi delle opportunità, delle sfide, dei rischi, dei problemi etici e delle criticità connesse all'utilizzo di sistemi di Intelligenza artificiale (IA) in ambito giuridico. Il libro, in linea con il livello generale del dibattito in materia, si rivela estremamente utile per acquisire elementi informativi finalizzati a comprendere un fenomeno in continua evoluzione, alcuni specifici aspetti tecnologici ed empirici e alcune prospettive di sviluppo, in merito al rapporto tra IA e diritto, con particolare riguardo all'impatto dei sistemi di IA nella sfera pubblica del 'decidere', nei contesti giuridici e amministrativi.

La domanda che sta alla base del lavoro si rinviene nelle pagine introduttive che sottolineano come la vera sfida consista nel comprendere se e in che modo stia cambiando il rapporto tra IA e decisione pubblica, soffermandosi su cosa renda attuale e relativamente innovativo il tema del "decidere con l'IA" (p. 9), alla luce dell'osservazione che l'espressione "Intelligenza artificiale" risulta essere stata coniata negli Stati Uniti nel 1956. Una domanda che deve, però, fare i conti con la realtà e con importanti sollecitazioni provenienti dal mondo della scienza, la più significativa delle quali si riscontra nell'osservazione finale: "il futuro dell'intelligenza artificiale rimane profondamente incerto, abbiamo tanti motivi di ottimismo quanti di preoccupazione" (p. 211).

Il volume si sviluppa in cinque capitoli strutturati per accompagnare il lettore in un percorso che parte da una prima ricostruzione, condotta con approccio realistico e descrittivo, circa lo stato dell'arte e gli sviluppi tecnologici che caratterizzano i sistemi di IA, specie i sistemi di IA generativa basati su reti neurali[1] e che connotano la relazione tra IA e diritto. Una relazione che si sviluppa in un ambito ampio e complesso che si estende dall'attività di giudici e avvocati (le fonti da cui attingono informazioni sui precedenti e sulle norme da applicare ai casi concreti, i mezzi tecnologici per estrarre ed elaborare tali informazioni, il modo di scrivere i rispettivi atti), fino all'uso dell'IA da parte degli apparati di supporto alle decisioni (come, ad esempio, le cancellerie dei tribunali, gli uffici giudiziari, le segreterie degli studi legali). Si passa quindi da un primo capitolo dedicato all'analisi dell'ingresso dell'IA nel diritto, considerato punto di arrivo di una dinamica millenaria che, da sempre, integra aspetti umani, sociali e tecnologici, ma che rappresenta anche una "nuova frontiera nel processo di compenetrazione tra diritto e tecnologie" (p. 20) in grado di investire tutti gli aspetti dell'attività giuridica. Al fine di cogliere le opportunità e i rischi dell'impiego delle tecnologie di IA nei diversi contesti professionali ed organizzativi, nello stesso capitolo se ne approfondisce l'esame con riferimento ai tre tipi di sistemi di IA, quelli esperti basati sulla conoscenza simbolica, quelli predittivi basati sull'apprendimento automatico e quelli generativi basati sui modelli linguistici.

Nei capitoli successivi, si forniscono elementi di analisi finalizzati a sviluppare modalità di utilizzo dell'IA che, attraverso informazioni mirate e corrette, possano coadiuvare il giudice nell'esercizio delle sue funzioni, supportandolo nei suoi compiti più complessi. In particolare, nel secondo capitolo si approfondisce l'intreccio tra algoritmi (intesi come una "sequenza di istruzioni che consentono di estrarre conoscenza dai dati", p. 81) e dati ("un insieme di valori che registrano informazioni", p. 83) usati per addestrare i modelli

1 Le reti neurali sono sistemi informatici che consistono di nodi (i cosiddetti neuroni, ovvero strutture computazionali che, ispirandosi al funzionamento del cervello umano, sono in grado di apprendere rappresentazioni complesse dei dati attraverso molteplici strati di elaborazione) collegati da connessioni (detti (dette?) anche parametri) alle quali sono assegnati pesi numerici. Per una ricostruzione di tali tecnologie si rimanda, tra gli altri, a Lovergine S., Occhiocupo G. (2024), *Elementi di analisi sull'impiego di sistemi e algoritmi di IA nelle decisioni amministrative*, Inapp Paper n.48, Roma, Inapp.

di apprendimento automatico o come input del modello per fare previsioni. Viene sottolineata l'importanza di prestare attenzione alla qualità, non solo tecnica ma anche etica, dei dati e degli insiemi di dati, ossia i dataset. Nello stesso capitolo viene fatto cenno a importanti questioni di natura giuridica sollevate dalla natura 'creativa' dell'IA, quali il diritto d'autore, la privacy e il modo in cui è intesa, nonché le relative modalità dell'anonimizzazione (in merito, viene espressa l'esigenza di giungere alla piena conoscibilità delle decisioni giudiziarie mediante banche dati ad hoc complete e accessibili), la separazione dei poteri che tenendo conto della digitalizzazione del giudicare promuova scelte tecniche basate su norme procedurali *digital by design* e la mancanza di una chiara visione della digitalizzazione delle attività giudiziarie. Il terzo capitolo è dedicato all'analisi dell'impatto delle applicazioni dell'IA sulle predizioni automatiche e sulle previsioni degli esiti delle controversie e alla loro connessione con le decisioni processuali. Il quarto capitolo si ricollega a quello precedente per ciò che attiene alle predizioni automatiche, pur soffermandosi sulla necessità della 'spiegabilità' di ogni decisione pubblica e della motivazione delle decisioni giurisdizionali, ritenuta una delle questioni più delicate nel raccordo tra uso dell'IA e principi giuridici. Il quinto capitolo affronta invece il tema di attualità della regolazione dell'IA, ovvero delle regole *per* l'IA, focalizzando però l'attenzione sulle regole che incidono sulle decisioni prese *con* l'IA. Attraverso la citazione dei principali documenti, rapporti e studi, viene delineata l'evoluzione culturale che ha portato, a partire dallo scorso decennio (2010-2020), all'affermazione in ambito europeo e internazionale dell'esigenza di regolamentare l'IA in una prospettiva etica (punto di arrivo di questo processo la Raccomandazione dell'Unesco del novembre 2021[2]). A livello internazionale, tale evoluzione ha portato all'adozione il 17 maggio 2024, da parte dei quarantasei membri del Consiglio d'Europa, della Convenzione quadro[3] contenente il primo trattato internazionale giuridicamente vincolante volto a garantire il rispetto delle norme giuridiche in materia di diritti umani, democrazia e Stato di diritto nell'utilizzo dei sistemi di IA. Il trattato, aperto anche ai Paesi non europei, stabilisce un quadro giuridico che copre l'intero ciclo di vita dei sistemi di IA, affronta i rischi che tali sistemi potrebbero presentare ed è diretto a tutelare principi fondamentali come la dignità umana, l'autonomia individuale, l'antropocentrismo, l'eguaglianza, la tutela dei dati personali, l'affidabilità basata su standard tecnici, l'integrità dei dati e la cybersicurezza. A livello europeo, viene citato il Regolamento (UE) n. 2024/1689 del 13 giugno 2024 – noto come *Artificial Intelligence Act (AI Act),* in vigore da agosto 2024 e che verrà pienamente applicato dal 2 agosto 2026[4] – volto all'istituzione di un quadro giuridico uniforme e nel quale si stabilisce che non vengano attribuiti compiti decisori a sistemi di IA. A livello nazionale, si fa cenno al DDL n. 1146 *Disposizioni e deleghe al Governo in materia di intelligenza artificiale*, attualmente all'esame finale delle Camere, che si pone l'obiettivo di dettare una disciplina che, senza sovrapporsi al Regolamento UE, predisponga un sistema di princìpi, governance e misure specifiche adatto al contesto italiano per cogliere tutte le opportunità dell'IA, contemplando, tra i princìpi e criteri direttivi, la formazione per i professionisti e gli operatori del settore, nonché l'accelerazione del processo di accrescimento delle competenze in ambito tecnologico all'interno dei percorsi di formazione e istruzione, nell'ottica di favorire l'inserimento dei giovani nel mondo del lavoro.

Una formazione, o meglio, un "piano di investimenti sul fattore umano" la cui necessità viene richiamata anche nelle pagine conclusive del libro, sulla base della consapevolezza che l'IA potrà modificare non solo il modo in cui apprendiamo, lavoriamo e interagiamo con gli altri, ma anche la più generale visione del significato stesso di essere persone che, con sempre maggiore frequenza, dovranno interagire e collaborare (e in alcuni casi competere) con macchine 'intelligenti', governandole, ma anche subendone l'influenza.

Alla luce del contesto tecnologico in evoluzione in cui operano i decisori pubblici, sia sul versante della PA che su quello giudiziario, e delle conseguenti implicazioni personali e sociali derivanti dalle loro decisioni con l'IA, gli Autori invitano a promuovere un'educazione mirata che parta dalle scuole (ove occorrono insegnanti con adeguate competenze e conoscenze da trasmettere ai giovani), prosegua nelle Università e si sviluppi con la

2 Unesco (2021), *Recommendation on the Ethics of Artificial Intelligence*, Paris, Unesco.

3 Council of Europe, Committee on Artificial Intelligence (CAI), *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, CM (2024) 52-final.

4 Per una sintetica analisi del Regolamento UE, si rimanda a Occhiocupo G., Lovergine S. (2024), *L'intelligenza artificiale nella pubblica amministrazione: opportunità e rischi alla luce dell'AI Act*, Futura Network, ASviS, https://oa.inapp.gov.it/handle/20.500.12916/4461.

formazione professionale di giudici e amministratori pubblici. In prospettiva, con riferimento al rapporto tra tecnologia, diritto e cultura, si suggerisce l'integrazione tra le competenze cognitive e pratiche dell'essere umano con quelle dell'IA, in modo da potenziare le proprie capacità di conoscere e agire, arrivando al cosiddetto 'decisore aumentato', onde evitare un indebolimento e una dequalificazione che porterebbe a una sostituzione dell'umano da parte dell'IA nella propria autonomia e dimensione lavorativa e sociale.
Dalla lettura del libro emerge l'ennesima conferma che l'IA rappresenti uno strumento utile, potente e rapidissimo, il cui livello di affidabilità dovrà essere costantemente verificato per mantenere e preservare la creatività e libertà "conservando la capacità tutta umana di sognare" (p. 211).

Giuditta Occhiocupo
INAPP

Per proporre un articolo

La Rivista pubblica articoli sui temi legati a monitoraggio, analisi e valutazione delle politiche del lavoro, dell'istruzione, della formazione, delle politiche sociali e, in generale, tutte le politiche economiche che hanno effetti sul mercato del lavoro.
Sinappsi pubblica solo articoli inediti. I contributi non possono pertanto riguardare articoli già pubblicati, anche solo in parte, su altre riviste italiane e straniere o essere contemporaneamente proposti ad altre riviste per la pubblicazione. I contributi possono essere proposti in lingua italiana o in lingua inglese. Gli articoli devono essere inviati in formato word all'indirizzo di posta elettronica sinappsi@inapp.gov.it.
I testi vanno corredati con la scheda di accompagnamento contenente la dichiarazione, sotto propria responsabilità, di originalità della proposta (https://www.inapp.gov.it/rivista-sinappsi/indicazioni-per-gli-autori) e dagli allegati (vedi infra).

Procedure

Ogni proposta, dopo la verifica della presenza dei requisiti minimi di pubblicabilità (rispetto delle norme editoriali) è soggetta all'esame preliminare del Comitato editoriale. Se giudicato coerente con gli obiettivi e gli standard qualitativi della Rivista, il testo è sottoposto, in forma anonima, al giudizio di almeno un referee in double blind peer review (ovvero autori e revisori reciprocamente anonimi). Il testo inviato ai revisori non deve pertanto contenere informazioni sull'identità degli autori. Questi sono quindi tenuti a minimizzare le autocitazioni e qualsiasi altro connotato che possa favorire la loro identificazione da parte dei revisori.

Il processo di revisione da parte dei referee potrà dar luogo a uno dei seguenti esiti: accettazione; accettazione subordinata a modifiche minori; accettazione subordinata a modifiche rilevanti; da sottomettere a riesame previe modifiche e senza impegno di successiva accettazione; rifiuto. L'accettazione subordinata a modifiche prevede la revisione da parte degli autori, che dovranno rendere evidenti nel testo le modifiche effettuate. Quando l'articolo è accettato per la pubblicazione, gli autori trasferiscono automaticamente all'Inapp ogni diritto di copyright, garantendo la possibilità della più ampia diffusione.

Agli autori sarà consegnata la prima bozza per la correzione, con l'invito a restituirla entro una data prefissata. Sulla prima bozza potranno essere apportate solo modifiche marginali. La correzione della seconda bozza sarà eseguita a cura della Redazione.

Dimensione e criteri di stesura dei testi:

- pagina formato A4;
- da 30.000 a 50.000 caratteri complessivi per articolo, spazi inclusi, comprese tabelle e figure (testi di dimensione superiore devono essere concordati con la Redazione, indicando i motivi per cui non è possibile rispettare i limiti previsti);
- titolo max 50 caratteri, eventuale sottotitolo max 70 caratteri, ma se è presente il sottotitolo, il titolo può avere un massimo di 30 caratteri;
- paragrafi numerati (solo primo livello); la Redazione può intervenire sul titolo proponendo modifiche agli autori;
- numero di tabelle + figure non superiore a 10 (non duplicare le informazioni fornite dalle tabelle con quelle dei grafici e del testo dell'articolo);
- tabelle e figure sempre numerate (ad esempio Tabella 1, Tabella 2 ecc.; Figura 1, Figura 2 ecc.), con titolo, fonte e anno. Possono essere utilizzati colori;
- note esplicative inserite a piè di pagina;

- richiami bibliografici inseriti nel testo entro parentesi tonde, con l'indicazione del cognome dell'autore da citare, seguito dalla data della pubblicazione originale ed eventualmente dalla/e pagina/e di riferimento della citazione riportata nel testo (Cognome autore data, numero pagina). A ogni richiamo deve corrispondere la fonte completa in bibliografia, inserita a fine saggio (vedi Norme bibliografiche);
- lo stesso sistema di rinvio alla bibliografia finale si adotta all'interno delle note a pie' di pagina.

Allegati

L'articolo va corredato da:

- una breve nota biografica (circa 600 caratteri spazi inclusi), elaborata in base al seguente modello: "Ricercatore/ trice, assegnista (oppure insegna) presso l'Istituto/Università (Denominazione). Aggiungere eventualmente, altri incarichi di prestigio. Fra le pubblicazioni recenti si segnalano: (indicare un max di due lavori). Indirizzo e-mail";
- dichiarazione, sotto propria responsabilità, di originalità della proposta (presente all'interno della scheda di accompagnamento);
- abstract in italiano e abstract in inglese di max 600 caratteri ciascuno, spazi inclusi;
- tre parole chiave in italiano e tre corrispondenti keyword in inglese;
- il file in formato Excel delle figure e dei grafici inseriti anche nel testo, un elemento per foglio, con numerazione corrispondente a quanto indicato nell'articolo.

Norme bibliografiche

Requisiti della bibliografia

- deve essere unica e collocata alla fine del lavoro;
- deve indicare esclusivamente le opere citate nel testo e nelle note ed essere aggiornata;
- deve prevedere l'ordine alfabetico per cognome dell'autore o del curatore, del primo autore o curatore nel caso di più nomi, e l'ordine cronologico di pubblicazione delle opere dalla più recente alla meno recente (per opere dello stesso autore pubblicate nello stesso anno, si usino le indicazioni a, b, c);
- i lavori di più autori vanno riportati con tutti i nomi.

Monografie

Autori:

Cognome autore e iniziali puntate del nome (anno tra parentesi), *Titolo del volume in corsivo. Se è presente, il sottotitolo va sempre in corsivo preceduto dal punto*, Luogo, Editore

Nel caso di più autori, mettere tutti gli autori separati da virgole.

Curatori:

Cognome curatore e iniziali puntate del nome (a cura di) (anno tra parentesi), *Titolo del volume in corsivo. Se è presente, il sottotitolo va sempre in corsivo preceduto dal punto*, Luogo, Editore

Per i testi stranieri mettere (eds.) al posto di (a cura di) nel caso di più curatori, (ed.) nel caso di curatore unico.

Nel caso di più curatori, mettere tutti i curatori separati da virgole.

Se la monografia fa parte di una collana, inserire nome della collana e relativo numero dopo il titolo.

Articoli di riviste/periodici

Cognome autore e iniziali puntate del nome (anno tra parentesi), Titolo dell'articolo in tondo. Se è presente, il sottotitolo va preceduto dal punto, Titolo del periodico/rivista in corsivo, annata[1], numero anno reso con n. e numero in cifre, pagine di inizio e fine articolo reso con pp. …-… (senza spazio dopo il punto. Es.: pp.33-45).

Nel caso di più autori, mettere tutti gli autori. Se presente, inserire il DOI tra parentesi uncinate < > senza spazi dopo e prima.

Estratti da monografie

Cognome autore e iniziali puntate del nome (anno tra parentesi), Titolo dell'estratto in tondo. Se è presente, il sottotitolo va preceduto dal punto, in Cognome autore e iniziali puntate del nome, *Titolo del volume in corsivo*, Luogo, Editore, pagine di inizio e fine articolo reso con pp. …-… (senza spazio dopo il punto. Es.: pp.40-60).

Nel caso di più autori, mettere tutti gli autori.

Se il volume di estrazione è a cura di, seguire le indicazioni per i volumi con curatore.

Testi Inapp

I testi Inapp seguono le indicazioni precedenti. Le monografie però devono SEMPRE riportare Inapp fra gli autori o curatori.

1 Annata: insieme di fascicoli di un periodico pubblicati nel corso di un anno o di un periodo editoriale determinato. Fonte http://elearning.unimib.it/mod/glossary/view.php?id=13076.

TESTI INAPP

Monografie

Inapp, Checcucci P., Fefè R., Scarpetti G. (a cura di) (2017), *Età e invecchiamento della forza lavoro nelle piccole e medie imprese italiane*, Inapp Report n.3, Roma, Inapp

Paper

Quaranta R., Ricci A. (2017), *Riforma delle pensioni e politiche di assunzione*. *Nuove evidenze empiriche, italiane*, Inapp Paper n.3, Roma, Inapp

Sinappsi

Cassese S. (2018), Evoluzione della normativa sulla trasparenza, *Sinappsi*, VIII, n.1, pp.5-7

Letteratura grigia

Schulz M., Bressers D., van der Steen M., van Twist M. (2015), Internal Advisory Systems in Different Political-Administrative Regimes, *Prepared for the International Conference on Public Policy (ICPP) T08P06 – Comparing policy advisory systems at the second International Conference on Public Policy, Milan 2015*

Comité de suivi du Cice, France Stratégie (2016), Comité de suivi du Crédit d'impôt pour la compétitivité et l'emploi. Rapport 2016, *Evaluation, Septembre* 2016

Giurisprudenza

Corte costituzionale

Corte cost. 25 luglio 1995 n.376, in *Giur. cost*., 1995, XL, 4, p.2750 ss.

Corte di Cassazione

Cass., sez. III, 14 ottobre 1991 n.10763, in *Dir. Trasp*., 1993, VI, 3, p.847 ss.

Cass. pen., sez. un., 26 marzo 2003, in *Cass. pen*., 2003, XLIII, 9, p.2579 ss.

Cass. pen., sez. VI., 3 novembre 2001, in *Riv. pen*. 2002, 1, p.31 ss.

Cass. civ., sez. lavoro, 29 maggio 1998, n.5348 Cass. pen., sez. I, 30 aprile 1992, Idda, in *C.E.D. Cass.* Pen., n.190564

Cass. pen., sez. un., 6 novembre 1992, Martin, in *Cass. Pen*., 1993, XXXIII, 2, p.280

Cass. pen., sez. IV, 21 ottobre 2005, in *Dir. Pen. Proc*., 2006, XII, 2, p.200

Cass. pen., sez. V, d 24 ottobre 2002, De Vecchis, in *Guida dir.*, 2003, X, 10, p.86

Consiglio di Stato

Cons. Stato, sez. IV, 14 giugno 2005 n.3120, in *Foro Amm. CDS*, 2005, IV, 6, p.1728 ss.

Corte dei conti

Corte conti 16 luglio 2010 n.15, in *Riv. corte conti,* 2012, LXV, 3-4, p.10

Corte d'Appello

App. Napoli 3 novembre 2008, in *Foro it*., 2009, CXXXIV, 5, pt. I, p.1476 ss.

Corte d'Assise

Corte Assise Milano 15 febbraio 2006, in *Giur. merito*, 2007, XXXIX, 3, p.783 ss., con nota di L.D. CERQUA

Tribunale

Trib. Roma 27 giugno 2005, in *Lavoro nella giur*., 2007, XV, 3, p.283, con nota di B. DE MOZZI

Tribunale Amministrativo Regionale

Tar Bari Puglia 6 aprile 2005 n.1376, in *Foro amm*.TAR, 2005, IV, 4, p.1214

Pretura

Pretore di Gubbio ord. 12 febbraio 1957, in Giur. cost., 1957, II, 1, p.127 ss.

Corte di Giustizia dell'Unione europea

Corte Giust., 28 giugno 1978, C-70/77, Simmenthal c. Amministrazione delle Finanze, in Racc., 1978, p.453

Letteratura grigia

La letteratura grigia segue le precedenti indicazioni rispetto al metodo Autore/Data. È necessario riportare sempre tutti gli elementi utili a rintracciare la pubblicazione:
Autori/Ente autore (anno), titolo del contributo, *informazioni aggiuntive*. Se disponibili, riportare il link al documento e/o il DOI tra parentesi uncinate < > senza spazi dopo e prima.

Giurisprudenza

Organo giurisdizionale emanante (Cassazione, Tribunale, Consiglio di Stato), tipo di atto adottato (Sentenza, Ordinanza, Decreto), sezione dell'organo emanante (non sempre presente), data della pronuncia, numero o nome delle parti (non sempre previsto. Se c'è il nome della parte dopo la data si tratta di un provvedimento della giurisdizione penale).

Legislazione

In ogni capitolo la prima citazione deve essere completa.
(Es. D.P.R. 26 luglio 1976 n.752, Norme di attuazione dello statuto speciale della Regione Trentino-Alto Adige in materia di ...)
Le citazioni successive possono essere in forma abbreviata.
(Es. D.P.R. n.752/1976)
La citazione degli articoli deve consentire l'individuazione precisa della disposizione normativa.
(Es. art. 5, comma 2, D.P.R. n.752/1976)
Citazioni all'interno del testo: legge n.150/2000, oppure L. n.150/2000

Risorse elettroniche

Le risorse elettroniche seguono le indicazioni precedenti rispetto al metodo Autore/Data. È SEMPRE necessario mettere il link al testo e/o pagina web di riferimento. Per le pagine inserire la dicitura (consultato il ….).

ESEMPI

Monografie

Campbell J.L., Pedersen O.K. (2014), *The national origins of policy ideas. Knowledge regimes in the United States, France, Germany and Denmark*, Princeton, Princeton University Press

Facchini C. (a cura di) (2008), *Conti aperti. Denaro, asimmetrie di coppie e solidarietà tra le generazioni,* Bologna, Il Mulino

Eichbaum C., Shaw R. (eds.) (2010), *Partisan Appointees and Public Servants, an International Analysis of the Role of the Political Adviser,* Cheltenham UK, Edward Elgar Publishing Limited

Eichhorst W., Wintermann O. (2005), *Generating Legitimacy for Labor Market and Welfare State Reforms. The Role of Policy Advice in Germany, the Netherlands and Sweden*, IZA Discussion Paper n.1845, Bonn, IZA

Articoli di riviste/periodici

Craft J., Halligan J. (2017), Assessing 30 years of Westminster policy advisory system experience, *Policy Sciences*, 50, n.1, pp.47-62 <DOI 10.1007/s11077-016-9256-y>

Estratti da monografie

Pattyn V., van Voorst S., Mastenbroek E., Dunlop C. A. (2017), Policy evaluation in Europe, in Ongaro E., Van Thiel S., *The Palgrave Handbook of Public Administration and Public Management*, Bristol, Policy Press, pp.105-11

Corte internazionale di Giustizia

Corte internazionale di Giustizia, sentenza del 27 giugno 1986, Attività militari e paramilitari contro il Nicaragua

Corte penale internazionale

Corte penale internazionale, Prima Camera di I grado, 14 marzo 2012, Thomas Lubanga Dyilo

Corte europea dei Diritti dell'Uomo

C. eur. Dir. Uomo, 12 febbraio 2013 – Ricorso n.24 818/03 – causa Armando Iannelli c. Italia

C. eur. Dir. Uomo, sentenza del 24 ottobre 1986, nel caso Agosi contro Regno Unito

Legislazione

- D.L. 27 giugno 1997 n.185
- D.M. 5 marzo 1999
- D.Lgs. 29 marzo 1993 n.119
- L. 13 febbraio 2001 n.45
- Art. 456 c. c.
- Art. 16, comma 4, lett. a, L. 28 gennaio 1994 n.84
- Art. 1 reg. CEE n.4056/86 del 22 dicembre 1986
- Regolamento n.1254/2008/CE della Commissione, che modifica il regolamento (CE) n.889/2008 recante modalità di applicazione del regolamento (CE) n.834/2007 del Consiglio relativo alla produzione biologica e all'etichettatura dei prodotti biologici, per quanto riguarda la produzione biologica, l'etichettatura e i controlli, in GU L 337 del 16.12.2008
- Direttiva n.70/50/CEE della Commissione, 22 dicembre 1969, in GUCE L 13, 19.1.1970


FSC
www.fsc.org
MISTO
Carta | A sostegno della gestione forestale responsabile
FSC® C103486

SINAPPSI è la rivista scientifica dell'Inapp (Istituto nazionale per l'analisi delle politiche pubbliche), ente di ricerca che svolge analisi, monitoraggio e valutazione delle politiche, in particolare di quelle che hanno effetti sul mercato del lavoro. Occupazione, istruzione, formazione e welfare sono i principali temi di studio. La rivista intende intrecciare, connettere appunto, i risultati e i contenuti della ricerca, per offrire al lettore una rete di riflessioni e ai decisori politici un impulso per le scelte strategiche.

ISSN 2532-8549

INAPP

ISBN: 978-88-543-0363-8