COESIONE ITALIA 21-27
GIOVANI, DONNE E LAVORO

Cofinanziato dall'Unione europea

MINISTERO del LAVORO e delle POLITICHE SOCIALI

FOCUS Inapp

INTELLIGENZA ARTIFICIALE IN 100 CONCETTI: UNA TASSONOMIA RAGIONATA DELL'ECOSISTEMA TECNOLOGICO

18

INAPP
ISTITUTO NAZIONALE PER L'ANALISI DELLE POLITICHE PUBBLICHE

ISSN: 3103-2788

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 e 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 (Piano Inapp 2023-2029), 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.

Riferimenti
Corso d'Italia, 33 00198 Roma
Tel. +39.06.85447.1
web: www.inapp.gov.it
Contatti: editoria@inapp.gov.it

Il lavoro è realizzato dal Gruppo di ricerca “Tecnologie e Intelligenza Artificiale: lavoro, professioni, formazione e competenze” della Struttura “Lavoro e Professioni". La pubblicazione è stata realizzata dall’Inapp in qualità di Organismo Intermedio del Programma nazionale Giovani, donne e lavoro FSE+ 2021-2027, Piano Inapp 2023-2029 - Operazione a titolarità n. 2 - Sviluppo del sistema Atlante del Lavoro e del Sistema informativo delle Professioni - Attività 6: Tecnologie e intelligenza artificiale: lavoro, professioni, formazione e competenze

Autori: Valentina Ferri, Boris Sofronic, Giuliana Tesauro

Testo pubblicato a marzo 2026

Impaginazione della collana a cura di Valentina Orienti
Elaborazione grafica copertina: Valentina Orienti

ISSN: 3103-2788

Indice

Introduzione

L'intelligenza artificiale (IA) è definibile come "una famiglia di tecnologie in rapida evoluzione" (Regolamento UE 2024/1689, AI Act), ossia un insieme di modelli e sistemi con differenti approcci operativi e finalità di applicazione.

L'utilizzo dell'IA certamente agevola la realizzazione di una pluralità di benefici nei diversi contesti di riferimento come quello economico, sociale o ambientale. Una sfida importante, tuttavia, è quella di contenere il verificarsi di eventuali rischi relativi all'utilizzo dell'IA e ai suoi livelli di sviluppo tecnologico, così da creare un contesto di fiducia e un clima di condivisione per una 'tecnologia antropocentrica' (Commissione Europea (2019) [1].

In tale ambito, l'alfabetizzazione[2] in materia di IA ricopre un ruolo centrale per massimizzare i benefici dei sistemi applicati e tutelare la sicurezza e i diritti fondamentali. Essa rappresenta un obbligo e un principio trasversale all'interno dell'AI Act dell'Unione europea.

In particolare, ai 'fornitori, ai *deployer*[3] e alle persone interessate' deve essere garantita la possibilità di ricevere le 'nozioni necessarie'[4] per assumere scelte informate sull'utilizzo dell'IA.

Si tratta di competenze che variano al variare del contesto di riferimento, ma che devono fornire una conformità e una corretta attuazione dei processi.

In questa prospettiva, attraverso interventi congiunti di alfabetizzazione e misure di follow-up, l'AI Act intende perseguire un miglioramento delle condizioni di lavoro e sostenere un percorso di innovazione affidabile, garantendo la promozione di strumenti di alfabetizzazione in materia di IA.

Nello specifico, è previsto che i fornitori e i *deployer* dei sistemi di IA adottino misure per assicurare "un livello sufficiente di alfabetizzazione" del proprio personale o di coloro che gestiscono tali sistemi per loro conto, considerando le diverse competenze ed esperienze (art.4, AI act).

L'AI *literacy*, dunque, prevede un approccio multidimensionale non solo tecnico, ma anche giuridico, etico e organizzativo. Si tratta di una vera e propria leva trasversale utile a rendere efficaci tutti gli strumenti del *risk-based approach* previsti dal regolamento.

Infatti, l'alfabetizzazione rappresenta un presupposto organizzativo e umano indispensabile per l'attuazione delle quattro dimensioni del *framework risk-based* dell'AI Act *(Human oversight, Risk management system, Data governance e System architecture*)[5] nel corso di tutto il ciclo di vita del sistema di IA, contribuendo all'incremento della compliance sostanziale.

---

[1] "I sette requisiti fondamentali che le applicazioni di AI dovrebbero soddisfare per essere considerate affidabili e antropocentriche sono: 1) intervento e sorveglianza umani; 2) robustezza tecnica e sicurezza; 3) riservatezza e governance dei dati; 4) trasparenza; 5) diversità, non discriminazione ed equità; 6) benessere sociale e ambientale; 7) accountability", in COM 2019/168 final.

[2] "Le competenze, le conoscenze e la comprensione che consentono ai fornitori, ai deployer e alle persone interessate, tenendo conto dei loro rispettivi diritti e obblighi nel contesto del presente regolamento, di procedere a una diffusione informata dei sistemi di AI, nonché di acquisire consapevolezza in merito alle opportunità e ai rischi dell'AI e ai possibili danni che essa può causare", Regolamento UE 2024/1689, Articolo 3 Definizioni, punto 56.

[3] "La nozione di deployer di cui al presente regolamento dovrebbe essere interpretata come qualsiasi persona fisica o giuridica, compresi un'autorità pubblica, un'agenzia o altro organismo, che utilizza un sistema di AI sotto la sua autorità, salvo nel caso in cui il sistema di AI sia utilizzato nel corso di un'attività personale non professionale. A seconda del tipo di sistema di AI, l'uso del sistema può interessare persone diverse dal deployer", Regolamento UE 2024/1689, considerando n.13.

[4] "Tali nozioni possono includere la comprensione della corretta applicazione degli elementi tecnici durante la fase di sviluppo del sistema di AI, le misure da applicare durante il suo utilizzo, le modalità adeguate a interpretare l'output del sistema di AI e, nel caso delle persone interessate, le conoscenze necessarie per comprendere in che modo le decisioni adottate con l'assistenza dell'AI incideranno su di esse", Regolamento UE 2024/1689, considerando n.13.

[5] Le quattro dimensioni trasversali del risk based approach dell'AI Act fanno riferimento ai seguenti articoli: Risk management system (art.9); Data governance (art.10); Human oversight (art.14); System architecture - Accuratezza, robustezza e cibersicurezza (art.15).

Partendo dall'insieme di tali considerazioni il presente lavoro intende sviluppare una tassonomia dell'intelligenza artificiale che, coerentemente con l'impianto dell'AI Act, chiarisca concetti essenziali dei sistemi di IA, riducendo le ambiguità terminologiche.

In questa prospettiva, la tassonomia non rappresenta un mero esercizio classificatorio, ma piuttosto uno strumento abilitante che da un lato rende effettivo l'obbligo di *literacy* normato dall'Unione europea e dall'altro partecipa all'acquisizione di competenze, conoscenze e capacità critiche sull'utilizzo dell'IA.

Il lavoro si pone un duplice scopo. Da un lato, intende contribuire al processo di divulgazione attraverso la sistematizzazione delle principali categorie e caratteristiche dei sistemi di intelligenza artificiale. Dall'altro, mira a costituire una base concettuale propedeutica a future ricerche multidimensionali, fornendo un linguaggio condiviso e una cornice interpretativa utile all'analisi degli impatti dell'IA sui sistemi produttivi e sul mercato del lavoro, senza alcuna pretesa di esaustività. In prospettiva, s'intende legare la tassonomia proposta proprio alle politiche di formazione delle competenze e delle conoscenze sia a livello istituzionale che organizzativo, in un'ottica di policy formative coerenti con l'impianto normativo europeo.

1. Tassonomie IA: rilevanza e contributi della letteratura

La velocità della diffusione dell'intelligenza artificiale (IA) in contesti eterogenei quali quello tecnologico, economico e sociale ha reso urgente l'esigenza di predisporre glossari e tassonomie condivise, al fine di offrire una terminologia comune alla vasta platea dei fruitori.

La creazione di definizioni comuni e classificazioni coerenti sostiene l'utilizzo di concetti chiari e condivisi, agevolando il percorso di regolamentazione e la comparabilità, rispetto alla valutazione degli impatti dell'IA.

Il compito ricoperto da una tassonomia specifica sull'IA, dunque, va oltre il mero aspetto descrittivo, poiché permette di filtrare principi etici come l'affidabilità, il rischio, l'uso e tradurli in categorie operative nei sistemi di IA.

Dall'analisi della letteratura emerge che le tassonomie sull'IA sono delle strutture concettuali multilivello utili ad organizzare domini di conoscenze a differenti livelli e non rappresentano un sistema unico. Esse si muovono prevalentemente su livelli di analisi distinti, rispecchiando la complessità dei sistemi di intelligenza artificiale. Le tassonomie, dunque, possono essere inquadrate come strumenti complementari tra loro che rispondono a esigenze tecniche, funzionali e settoriali.

Una ricerca fondamentale in questo campo è rappresentata dal contributo di Newman (2023) che propone una tassonomia dell'affidabilità per l'intelligenza artificiale a supporto di un AI *Risk Management Framework* (RMF) sviluppato dal NIST - *National Institute of Standards and Technology*. Nel caso specifico il livello della tassonomia adottato è orientato a rischi, impatti e affidabilità (effetti e vulnerabilità dell'IA). In particolare, ci si sofferma sugli effetti potenzialmente negativi generati dall'IA su individui o gruppi sociali e istituzioni, ponendo l'attenzione su aspetti fondamentali nei processi di valutazione, auditing e gestione del rischio.

La tassonomia proposta individua 150 proprietà che contribuiscono alla 'fiducia' in un sistema di IA.

Le proprietà sono mappate sui seguenti aspetti:

- una caratteristica di *trustworthiness* (come definita dal NIST);
- una fase del ciclo di vita dell'IA;
- sezioni specifiche dell'IA RMF nelle quali individuare strumenti/guida per implementare una proprietà.

In particolare, le caratteristiche di affidabilità (*trustworthy*) definite nel NIST AI RMF sono le seguenti:

- validità e affidabilità;
- sicurezza;
- resilienza;
- responsabilità e trasparenza;
- spiegabilità e interpretabilità;
- privacy;
- equità e gestione dei bias dannosi.

A queste si aggiunge una ulteriore caratteristica *Responsible Practice and Use* (Pratiche responsabili di sviluppo e uso) a sottolineare la necessità di sviluppare una fiducia non solo verso aspetti tecnici, ma anche sulle dinamiche e i percorsi attraverso i quali i sistemi di IA vengono sviluppati e gestiti in campo privato e sociale.

Tali caratteristiche sono poi utilizzate per costruire 'famiglie concettuali' al fine di strutturare le proprietà di *trustworthiness*.

Un ulteriore passo distingue sette fasi del ciclo di vita dell'IA:

- pianificazione e progettazione;
- raccolta e trattamento dei dati;

- costruzione e uso del modello;
- verifica e validazione;
- distribuzione e utilizzo;
- operazioni e monitoraggio;
- misurazione dell'impatto sull'uso di una IA da parte degli utenti.

Per ciascuna fase del ciclo di vita, la tassonomia prevede proprietà da considerare e collegamenti alle pratiche di gestione del rischio, al fine di garantire in modo sistematico la fiducia nel percorso di sviluppo, utilizzo e monitoraggio dell'IA.

Parallelamente, una recente ricerca (Abercrombie *et al*. 2024) ha esteso il focus dalle caratteristiche dei sistemi di IA agli impatti e potenziali danni. Gli autori propongono una tassonomia collaborativa e *human-centred* dei danni (*harms*) derivanti da IA, algoritmi e automazione. Si tratta di un approccio *human-centred* e orientato all'uso pensato per essere accessibile anche ai non specialisti, basandosi su esperienze e danni subiti dalle persone o dai contesti sociali. Il punto di partenza della ricerca è rappresentato dalle attività umane, dagli obiettivi e dalle interazioni uomo-IA, indipendentemente dalle specifiche implementazioni tecnologiche. La tassonomia in oggetto organizza i danni su due livelli principali:

1. tipo di danno (*harm type*) che esprime una categoria generale degli effetti negativi (danno fisico o materiale, danno psicologico o sociale, violazioni di diritti o libertà, impatto economico negativo, distruzione di fiducia o reputazione);
2. sottocategorie di danno che sono più specifiche e descrivono le modalità di manifestazione del danno.

L'intero percorso tassonomico è concepito nel rispetto di specifici principi: collaborativo, *human-centred*, flessibile ed estensibile, interoperabile.

Tale approccio completa le tassonomie orientate alla *trustworthiness*, spostando l'attenzione dagli attributi tecnici ai risultati concreti per individui e comunità.

Altri contributi si soffermano sulla classificazione delle tipologie di IA, come nel caso dello studio di Strobel *et al*. (2024) che indaga empiricamente le applicazioni reali della *Generative Artificial Intelligence* (GenAI) per approfondirne caratteristiche e natura. In questo caso si propone una tassonomia che si caratterizza per il livello funzionale, spiegando cosa fa l'IA (*Generator, Reimaginator, Synthesizer* etc.) in base a compiti e funzioni, per tipo di output e trasformazione che produce. In particolare, gli autori studiano cento casi reali di applicazioni GenAI e propongono una tassonomia dettagliata che identifica dimensioni e caratteristiche chiave per descrivere e differenziare applicazioni GenAI.

Gli autori si basano su di un framework iniziale e su cinque tipi di GenAI ciascuno con una funzione centrale:

- Generator (crea contenuti ex novo – testo, immagini);
- Reimaginator (trasforma contenuti esistenti cambiandone forma, stile o significato, ad esempio: riscrittura di un testo in un altro stile);
- Synthesizer (combina più input dati, testi, fonti per produrre un output unico);
- Assistant (supporta l'utente nello svolgimento di compiti);
- Enabler (non è usato direttamente dall'utente finale).

Le cinque tipologie derivano dalle caratteristiche fondamentali[6] del 'nuovo paradigma tecnologico' rappresentato dalla GenAI.

[6] Caratteristiche fondamentali della GAI sono: capacità generativa, probabilistica, apprendimento da grandi dataset, multimodalità, adattabilità all'input umano.

La struttura tassonomica comprende:

- 3 meta-dimensioni (livelli di analisi generici che guidano la classificazione);
- 10 dimensioni (aspetti specifici utili per la valutazione di ciascuna applicazione come input/output, capacità tecniche, modelli di business etc.);
- 38 caratteristiche (proprietà concrete degli aspetti tecnici e funzionali delle applicazioni GenAI).

Un ulteriore approccio riguarda le tassonomie orientate all'uso e al contesto socioeconomico dell'IA. In Theofanos *et al*. (2024) si adotta un approccio *human-centered* per classificare gli usi dell'IA in relazione alle crescenti interazioni uomo-IA. Gli autori pongono particolare attenzione alla classificazione delle modalità con le quali un sistema di IA contribuisce a un risultato. La tassonomia proposta individua 16 'attività'[7] di utilizzo dell'IA, indipendenti dai metodi di implementazione, ciascuna delle quali descrive un contributo specifico dell'IA alle attività umane; tali attività possono essere combinate per modellare compiti più complessi. Anche in questo caso si tratta di una tassonomia funzionale, dal momento che classifica l'IA in base agli usi concreti e alle tipologie di applicazioni sui processi organizzativi e umani.

A livello settoriale e macroeconomico l'OECD (2024) introduce una *sectoral taxonomy of AI intensity*, che consente di misurare e confrontare il grado di penetrazione dell'IA nei diversi settori produttivi, valutando l'integrazione dell'IA nell'economia. Questo approccio facilita il confronto tra il grado di integrazione dell'IA nei sistemi produttivi, rispetto al capitale umano, all'innovazione o al grado di esposizione alla tecnologia.

Lo studio succitato propone una tassonomia settoriale dell'intensità dell'IA, individuando diverse dimensioni in relazione alle attività dei differenti settori economici.

L'intensità dell'IA è una misura multidimensionale che riflette quanto l'IA influisce su un settore in termini di:

- capitale umano (AI *human capital*);
- innovazione (AI *innovation*);
- esposizione potenziale all'IA (*barrier-adjusted* AI *exposure*);
- uso effettivo dell'IA (AI *use*)[8]

Ogni settore economico riceve una valutazione su ciascuna dimensione, così da ottenere non un punteggio assoluto, ma un insieme di valori sui differenti profili[9]. Si propone, infine, un indicatore sintetico che consente un confronto settoriale.

---

[7] Di seguito l'elenco delle 16 attività di uso dell'IA definite dalla AI Use Taxonomy: A Human-Centered Approach del NIST: 1 Content Creation - generazione di nuovi artefatti come testo, immagini, codice o dati sintetici. 2. Content Synthesis - combinazione e sintesi di parti o concetti in un tutto coerente. 3. Decision Making -selezione di un corso d'azione tra alternative per arrivare a una soluzione. 4. Detection - individuazione o identificazione di qualcosa tramite ricerca o esame. 5. Digital Assistance - assistenza personale conversazionale o esecutiva su comandi e richieste. 6. Discovery -scoperta o individuazione di nuove informazioni o relazioni. 7. Image Analysis - riconoscimento di attributi o caratteristiche in immagini digitali. 8. Information Retrieval/Search - ricerca e recupero di informazioni su argomenti specifici. 9. Monitoring - osservazione continuata di processi o stati per ottenere informazioni sull'andamento. 10. Performance Improvement - miglioramento della qualità o dell'efficienza dei risultati. 11. Personalization - adattamento di un prodotto, servizio o contenuto alle caratteristiche o preferenze individuali. 12. Prediction -anticipazione di eventi o risultati futuri. 13. Process Automation - automazione di processi o attività ripetitive. 14. Recommendation - suggerimento di scelte o contenuti pertinenti all'utente. 15. Robotic Automation - controllo automatizzato di sistemi robotici. 16. Vehicular Automation - automazione di veicoli o sistemi di trasporto.

[8] 1) Capitale umano per l'IA (AI human capital) misura la domanda di competenze in IA nel mercato del lavoro (percentuale di annunci di lavoro con richieste di competenze AI); 2) Innovazione in IA (AI Innovation) valuta quanto un settore produce innovazioni legate all'IA (numero o quota di brevetti collegati all'AI sviluppati dal settore); 3) Esposizione potenziale all'IA (Barrier-adjusted AI Exposure) mostra quanto i compiti e le attività del settore sono potenzialmente influenzabili dall'IA, tenendo conto di barriere reali (come costi, scarsa maturità delle tecnologie, mancanza di competenze, regolamentazioni etiche ecc.); 4) Uso attuale dell'IA (AI use) rappresenta l'adozione effettiva dell'IA nelle imprese (indagini statistiche sulla percentuale di imprese che dichiarano di usare tecnologie IA).

[9] "While some sectors, such as IT services, score high along all the dimensions considered, others, such as Pharmaceuticals, exhibit more considerable heterogeneity (high AI human capital but low AI innovation)", in OECD (2024), A Sectoral Taxonomy of AI Intensity, OECD Artificial Intelligence, Papers December 2024 No. 30, p.3.

Infine, alcune tassonomie si concentrano sull'ecosistema e sui modelli di business dell'IA[10]. Il lavoro di Paeplow *et al*. (2025) su AI *startups for good* propone classificazioni dei modelli di business sostenibili basati sull'IA, evidenziando il legame tra tecnologia, valore sociale e sostenibilità ambientale, soffermandosi su configurazioni organizzative e creazione di valore. In primo luogo, lo studio in oggetto propone una tassonomia dei modelli di business delle startup basate sull'intelligenza artificiale, muovendosi sull'analisi di cento casi reali, al fine di 'catturare il modo in cui l'intelligenza artificiale viene attualmente utilizzata nelle startup per raggiungere un valore di sostenibilità'[11]. Successivamente, si presentano cinque configurazioni archetipiche di modelli di business:

- AI environmental analyser;
- AI healthcare improver basato su dati dei pazienti;
- AI product manufacturer per l'agricoltura e il grocery;
- AI surveillant and reporter di dati forniti dai clienti;
- AI energy improver.

Tali archetipi riflettono le possibili combinazioni tra le proprietà comuni dei modelli di business.

Sullo stesso filone tassonomico, il report dell'EIT - *European Institute of Innovation & Technology* - (2021) presenta una tassonomia che struttura l'ecosistema dell'AI su più dimensioni complementari, "ciascuna delle quali cattura un aspetto chiave del funzionamento e dell'impatto dell'intelligenza artificiale". In particolare, riconosce varie dimensioni:

- dimensione tecnologica (le principali tecniche e capacità dell'AI ad esempio apprendimento automatico, analisi dei dati etc. – per comprendere che tipo di AI viene utilizzata);
- dimensione applicativa (i settori e gli ambiti di utilizzo dell'IA es: sanità, industria, energia etc. per evidenziare dove l'IA viene applicata);
- dimensione degli attori (startup, grandi imprese, centri di ricerca, università e istituzioni pubbliche, per capire chi contribuisce allo sviluppo dell'ecosistema);
- dimensione del valore e dell'impatto (benefici economici, sociali e ambientali generati dall'IA, con particolare attenzione agli obiettivi di sostenibilità e interesse pubblico).

Accanto alle classificazioni concettuali e tassonomiche dell'intelligenza artificiale, la letteratura scientifica evidenzia l'importanza di definizioni condivise e operative dei termini. In questo quadro, i glossari assumono un ruolo complementare alle tassonomie, chiarendo i concetti e riducendo le ambiguità terminologiche. La rapida evoluzione delle tecnologie di IA rende infatti indispensabile un linguaggio comune, soprattutto nei contesti interdisciplinari e applicativi, per garantire una comprensione coerente e un uso consapevole dei concetti chiave.

In tal senso, Estévez Almenzar *et al*. (2022) suggeriscono un glossario di intelligenza artificiale *human-centric*, sempre orientato all'uso, che fa riferimento a una IA affidabile in termini di trasparenza, responsabilità o equità. La raccolta si basa su 230 termini derivati da oltre 10 diverse fonti generali (standard, documenti politici e testi giuridici, riferimenti scientifici). Ogni termine è corredato da una o più definizioni collegate ai riferimenti e integrato con le definizioni degli autori qualora privo di definizioni ad hoc.

[10] "Business model taxonomies can have multiple purposes, such as understanding, classifying, or designing business models, for which reason they pose a valuable artifact for this study purpose for structuring AI startups on sustainability as well", in Paeplow J. *et al* (2025), AI Startups for Good: A Taxonomy and Archetypes of Sustainable Business Models, Journal of Cleaner Production, 520, 146144.

[11] "To build the taxonomy, we follow the method by Kundisch D. *et al*. (2022), which follows a design-oriented paradigm",I n Paeplow J. *et al* (2025), AI Startups for Good: A Taxonomy and Archetypes of Sustainable Business Models, Journal of Cleaner Production, 520, 146144.

Più di recente, l'attenzione si è spostata verso l'intelligenza artificiale generativa Galindo-Cuesta (2025) che sviluppa un glossario concettuale orientato al contesto educativo intrecciando i concetti tecnici dell'IA e le pratiche pedagogiche. Il glossario applica un approccio misto tra revisione sistematica della letteratura (seguendo le linee guida PRISMA - *Preferred Reporting Items for Systematic Reviews and Meta-Analyses*), tecniche di elaborazione del linguaggio naturale (analisi della frequenza dei termini, clustering semantico etc.) e un processo di validazione Delphi con esperti multidisciplinari. La selezione dei termini avviene in base alla "rilevanza pedagogica, alla chiarezza concettuale e alla frequenza d'uso nel discorso sull'IA in ambito educativo". Si presentano, in tal modo, definizioni operative dei termini chiave associati all'IA generativa, in particolare i modelli linguistici di grandi dimensioni (LLM).

Analogamente, Cherner *et al*. (2025) elaborano un glossario essenziale *human-centric* e orientato all'uso per l'IA generativa nella *higher education*, evidenziando il ruolo dei glossari come strumenti di mediazione concettuale tra innovazione tecnologica e pratiche didattiche. Gli autori hanno realizzato un glossario fondamentale di 24 termini necessari per l'utilizzo della GenAI negli ambienti dell'istruzione superiore, insieme ad esempi applicativi per ciascun termine.

Inoltre, sono disponibili on-line *repository* terminologici che contengono risorse di riferimento per la standardizzazione terminologica nell'IA, offrendo definizioni affidabili, aggiornate e categorizzate dei concetti chiave (Google Developers, Stanford University, MIT Media Lab, Agenda Digitale).[12]

In sintesi, attraverso approcci incrociati e complementari, tassonomie e glossari favoriscono l'allineamento semantico tra ricerca, sviluppo, regolazione e uso dell'intelligenza artificiale, supportando processi decisionali informati e una governance coerente dei sistemi intelligenti.

L'impostazione multilivello delle tassonomie, pertanto, non rappresenta una frammentazione dell'impianto concettuale, ma piuttosto ne dilata la comprensione in un contesto variegato di applicazioni e sistemi in veloce evoluzione. Nessuna lettura, quindi, può ritenersi esaustiva isolatamente, ma va integrata con i differenti livelli per una lettura chiara dei fenomeni in atto e degli effetti che essi producono.

[12] Quarteroni A., Regazzoni F. (2022), *Un glossario per l'intelligenza artificiale: da Algoritmo a Unsupervised Learning*, https://www.agendadigitale.eu/cultura-digitale/un-glossario-per-lintelligenza-artificiale-da-algoritmo-a-unsupervised-learning/

2. Metodologia e criteri di selezione dei termini

La tassonomia proposta non si configura come un glossario o come un catalogo di tecnologie, si tratta infatti di una mappa sistemica dell'ecosistema dell'intelligenza artificiale. L'obiettivo è di ricostruire l'architettura multilivello che rende possibile l'IA contemporanea, distinguendo tra i fondamenti concettuali anzitutto cosa è l'IA e quali nozioni la strutturano. In secondo luogo, vengono approfondite le famiglie tecnologiche e quindi quali approcci e logiche operative la caratterizzano. Successivamente si passano in rassegna i modelli e le applicazioni; gli attori industriali e l'infrastruttura (chi la sviluppa e su quali basi hardware e software). Un'altra parte rilevante riguarda i metodi di apprendimento e valutazione. Infine, si passa alla dimensione regolatoria ed etica; i rischi e i limiti.

In tal senso la costruzione data a tale classificazione riflette l'idea che l'intelligenza artificiale sia un fenomeno sociotecnico complesso, in cui tecnologia, mercato, governance e cultura risultano strettamente interconnessi. L'inclusione di modelli, aziende, prodotti, infrastrutture e concetti critici non risponde, pertanto, a una logica enciclopedica, ma a un criterio analitico: rappresentare l'IA come ecosistema stratificato, necessario per comprendere le sue implicazioni su lavoro, competenze, istituzioni e dinamiche economiche.

La tassonomia proposta risponde all'esigenza di individuare una struttura concettuale unificata, capace di integrare aspetti tecnici, funzionali, socioeconomici e regolatori dell'intelligenza artificiale in un quadro coerente. Un primo obiettivo è la definizione di un lessico condiviso che riduca le ambiguità terminologiche, spesso ricorrenti nel dibattito pubblico e scientifico. A tal fine, per ciascun concetto è indicata una fonte principale di riferimento; tuttavia, in alcune circostanze, la definizione viene ampliata rispetto alla fonte citata, adottando un linguaggio accessibile e ricorrendo, ove opportuno, a esempi e metafore esplicative, al fine di rendere più chiaro il significato dei termini e facilitarne la comprensione anche a un pubblico non strettamente specialistico.

Dal momento che l'obiettivo è andare oltre il puro aspetto descrittivo di un glossario, i concetti fondamentali dell'IA vengono collegati alle implicazioni applicative, ai rischi, ai requisiti di governance per forgiare uno strumento operativo utile all'AI *literacy* prevista dall'AI Act.
Il lavoro è quindi un impianto tassonomico trasversale e interdisciplinare in linea con la letteratura esistente e ad essa complementare, in quanto coniuga in un'unica struttura concettuale prospettive tradizionalmente separate e mira a favorire coerenza concettuale e chiarezza interpretativa, senza alcuna pretesa di esaustività. Si riprendono, infatti, principi orientati alla *trustworthiness* e al rischio (NIST AI RMF, Newman 2023), alle tassonomie *human-centred* dei danni (Abercrombie *et al*. 2024), agli approcci tecnico-strutturali tipici dell'informatica (Strobel *et al*. 2024, Wilson 2023). Tali prospettive sono rese interoperabili attraverso un glossario essenziale che lega descrizioni tecniche e applicazioni pratiche.

La struttura multilivello della tassonomia si basa sulla selezione di categorie emerse da un'analisi comparativa delle principali tassonomie tecniche e dei framework regolatori europei, individuando convergenze concettuali e terminologiche. Pur potendo essere consultato in modo non lineare, il lavoro presenta numerose interconnessioni tra i termini, che rendono la struttura tassonomica trasversale e, per certi aspetti, parallela nei livelli di approfondimento. Per questo motivo si suggerisce una lettura completa del testo che consente di acquisire progressivamente i concetti di base e di comprendere quelli più avanzati, evitando fraintendimenti sui termini essenziali, talvolta di uso frequente ma non sempre chiari nel loro significato.

La selezione dei termini è avvenuta attraverso una ricognizione della letteratura scientifica relativa agli anni più recenti, dei principali framework tecnici (NIST AI RMF, ISO/IEC JTC 1/SC 42) e sulla base dei documenti regolatori europei (AI Act, linee guida della Commissione Europea), nonché tenendo conto delle tassonomie tecnologiche elaborate da istituzioni accademiche e centri di ricerca internazionali.

Sono stati inclusi i lemmi che soddisfacevano uno o più dei requisiti di seguito riportati: la rilevanza strutturale nel funzionamento dei sistemi di IA; la centralità nel dibattito regolatorio o etico; la diffusione consolidata

nella pratica applicativa o industriale o organizzativa nonché l'impatto significativo sulle dinamiche del lavoro, delle competenze e dei modelli aziendali.

Ne emerge una tassonomia che integra aspetti tecnici, sociali e applicativi, riconoscendo l'IA come un fenomeno complesso che coinvolge tecnologie, persone e contesti organizzativi.
Tale tassonomia può rappresentare una base teorica per la mappatura delle skill IA attraverso l'analisi della trasformazione del lavoro e lo studio dei rischi di sostituzione e ibridazione cognitiva (Ferri *et al*. 2024, Ferri *et al*. 2025 (a), Ferri *et al*. 2025 (b), Marsiglia *et al*. 2025). In sintesi, in presenza di una IA matura, tecnologia e competenza diventano strutturalmente integrate. La base di tale approccio consiste nel considerare l'IA come *General Purpose Technology* secondo l'impostazione di Bresnahan e Trajtenberg (1995), caratterizzata da pervasività (diffusione trasversale a settori e funzioni), miglioramenti continui e complementarità con innovazioni organizzative e competenze. Le tecnologie ridefiniscono la struttura del lavoro, consentendo "al capitale di sostituire il lavoro in compiti in cui era precedentemente impegnato" (Acemoglu e Restrepo 2019). Tale impatto si concretizza non solo in sostituzione capitale-lavoro (che non necessariamente riduce la domanda aggregata di lavoro), ma anche in complementarità tra task che innescano risposte compensative sulla produzione (settoriale e intersettoriale) e sulla domanda finale (Autor e Salomons 2018). L'intero lavoro cognitivo subisce un processo di riconfigurazione accompagnato da incrementi di produttività e differenze sostanziali tra i vari operatori (Brynjolfsson *et al*. 2025). Sui differenti e potenziali impatti di modelli linguistici sulle mansioni lavorative si concentra un framework proposto da Eloundou *et al.* (2024).
Ne deriva una configurazione dell'IA non solo come tecnologia esterna, ma anche come infrastruttura cognitiva incorporata nelle competenze. Pertanto, la sovrapposizione tra tecnologia e skill nei *job posting*, la presenza simultanea di IA come tecnologia e come skill nei dati non è ridondanza semantica ma costituisce evidenza empirica della sua natura generale e trasversale.

La tassonomia proposta può quindi costituire una base teorico-empirica per la mappatura evolutiva delle competenze IA e per lo studio dei processi di sostituzione e ibridazione cognitiva, contribuendo alla riconfigurazione dell'architettura cognitiva del lavoro.

3. Fondamenti e tipologia di IA

3.1 Concetti fondamentali dell'intelligenza artificiale

In questa sezione sono raccolti termini e concetti che costituiscono le nozioni essenziali senza le quali sarebbe difficile orientarsi nel panorama sempre più articolato delle tecnologie intelligenti. Nel paragrafo si trova la definizione stessa di intelligenza artificiale, declinata sia dal punto di vista della capacità tecnologica sia dal punto di vista del ruolo di disciplina scientifica, insieme ai concetti che ne descrivono le ambizioni più elevate – dall'intelligenza artificiale generale (AGI) alla singolarità tecnologica. Viene inoltre introdotto il *machine learning*, inteso come il principale paradigma operativo attraverso cui i sistemi di IA apprendono dai dati e migliorano le proprie prestazioni nel tempo. Si tratta di nozioni che aiutano a contestualizzare il presente tecnologico e che possono essere d'aiuto nella comprensione del potenziale dello strumento.

La sezione approfondisce inoltre gli elementi concettuali e le unità più elementari su cui si regge ogni sistema di intelligenza artificiale: gli algoritmi come sequenze di istruzioni che governano il funzionamento delle macchine, i Dataset come raccolte di dati da cui i modelli apprendono, i Big Data come fenomeno quantitativo che alimenta le capacità predittive dell'IA e i Token come unità minime di elaborazione del linguaggio. È possibile poi comprendere anche come funziona l'inferenza, ovvero il momento in cui un modello addestrato viene effettivamente utilizzato per produrre risultati e quali strumenti possono essere utili nella interpretazione di video e immagini.

Infine, vengono illustrati strumenti di interazione quotidiana come i chatbot, che hanno reso l'IA maggiormente accessibile attraverso il dialogo in linguaggio naturale. Al termine della lettura di questa sezione si disporrà di una prima parte di vocabolario solido per affrontare con consapevolezza le sezioni successive della tassonomia.

AGI (Artificial General Intelligence)

L'intelligenza artificiale generale (AGI) indica l'ipotesi di un sistema artificiale capace di svolgere molti tipi diversi di compiti cognitivi con una flessibilità paragonabile a quella umana, anziché essere limitato a un singolo ambito.
Si tratta di una nozione teorica e prospettica: allo stato attuale non esistono sistemi riconducibili a tale definizione. Il concetto è oggetto di ampio dibattito scientifico, tecnologico e industriale e non esiste una definizione condivisa né un consenso circa la possibilità e le modalità di un suo eventuale sviluppo.

Fonte: Bostrom N., (2014), Superintelligence: Paths, Dangers, Strategies, Oxford University Press; cfr. OpenAI Charter: openai.com/charter/

AI Inference

Si tratta della fase operativa di un modello di intelligenza artificiale: dopo l'addestramento, il sistema utilizza i parametri appresi per generare output su dati nuovi che non erano presenti nei dati di training. È la fase in cui il sistema viene effettivamente utilizzato in contesti reali, con o senza interazione diretta dell'utente.

Fonte: https://www.ibm.com/cloud/learn/ai-inference

Algoritmo

Un algoritmo è una procedura computazionale ben definita costituita da una sequenza ordinata di passaggi che trasformano uno o più valori in input in uno o più valori in output. Può essere inteso come uno strumento per risolvere un problema computazionale specificato, realizzando una determinata relazione tra input e output attraverso operazioni formalizzate.
In altre parole, è una procedura che indica passo dopo passo come risolvere un determinato problema, applicando regole precise e definite.

Fonte: Cormen T. H., Leiserson C. E., Rivest R. L., Stein C. (2022), Introduction to algorithms, The MIT Press

Big Data

I Big Data sono grandi e complesse quantità di dati che non possono essere gestite con strumenti tradizionali. Vengono raccolti da fonti come social media, sensori, transazioni online e dispositivi connessi. I Big Data sono comunemente descritti attraverso le '3V': volume (quantità), varietà (diversità di formati e fonti) e velocità (rapidità di generazione e aggiornamento), a cui spesso si aggiungono veridicità (qualità e affidabilità) e valore (utilità informativa).
I sistemi di intelligenza artificiale possono utilizzare grandi quantità di dati per migliorare le proprie prestazioni, ma la qualità e la pertinenza dei dati restano fattori decisivi.

Fonte: Cambridge Dictionary: Big Data
Laney D. (2001), 3D Data Management: Controlling Data Volume, Velocity, and Variety, Gartner Research Note

Chatbot

Il chatbot è un programma informatico progettato per simulare una conversazione con un essere umano, attraverso testo o voce. I chatbot più semplici seguono regole predefinite, mentre quelli basati sull'IA (come, per esempio, ChatGPT o Claude) sono in grado di elaborare domande complesse e generare risposte articolate in linguaggio naturale.

Fonte: Oracle Glossary: Chatbot

Computer Vision

Il campo dell'intelligenza artificiale che sviluppa metodi e modelli per consentire ai sistemi informatici di analizzare e interpretare immagini e video. Le applicazioni sono innumerevoli: dal riconoscimento facciale per sbloccare lo smartphone, al controllo qualità nelle fabbriche, fino alla guida autonoma delle automobili.

Fonte: Szeliski R. (2022), Computer Vision: Algorithms and Applications. 2nd edition, Springer Nature

Dataset (in ambito IA)

Il dataset è una raccolta strutturata di dati utilizzata per l'addestramento, la validazione o la valutazione di un modello di intelligenza artificiale. Un dataset può comprendere diverse tipologie di dati (testo, immagini, dati numerici, audio o video).
I dataset utilizzati nei sistemi di intelligenza artificiale possono risultare disallineati rispetto al contesto originario o diventare obsoleti rispetto al contesto di applicazione. Particolare attenzione è richiesta quando i modelli sono addestrati su grandi quantità di dati sensibili o quando i loro output producono effetti diretti o indiretti sulle persone. La qualità e la quantità dei dati contenuti nel dataset influenzano le prestazioni del modello. L'affidabilità (*trustworthiness*) dei sistemi di IA dipende non solo da modelli e algoritmi, ma anche dalla qualità e rappresentatività dei dataset, dai contesti organizzativi e dalle decisioni dei progettisti e dei supervisori umani. *Bias* computazionali e statistici possono emergere da campioni non rappresentativi o da errori sistematici nei dati.

Fonte: Tabassi E. (2023), Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST Trustworthy and Responsible AI, National Institute of Standards and Technology, Gaithersburg, MD, [online], https://doi.org/10.6028/NIST.AI.100-1

Intelligenza Artificiale (IA)

È l'insieme di metodi e tecniche computazionali progettati per eseguire compiti che, nei contesti umani, sono associati a capacità cognitive quali l'elaborazione del linguaggio, il riconoscimento di pattern o il supporto decisionale. È al tempo stesso una disciplina scientifica che studia e sviluppa tali metodi e l'insieme dei sistemi e delle applicazioni che ne derivano.
Le sue applicazioni spaziano dai servizi digitali di uso quotidiano – come i sistemi di raccomandazione, gli assistenti vocali, i filtri antispam e la traduzione automatica – fino a sistemi avanzati integrati nei processi produttivi, nella sanità, nella finanza e nella pubblica amministrazione. In questi contesti, l'adozione dell'IA richiede infrastrutture tecnologiche adeguate, competenze specialistiche e investimenti organizzativi mirati.

Fonte: Britannica: https://www.britannica.com/technology/artificial-intelligence
Enciclopedia Treccani: Intelligenza Artificiale

IoT (Internet of Things) e AIot

Il termine *Internet of Things* è utilizzato per descrivere dispositivi integrati, le 'cose', dotati di connettività a Internet, che consente di interagire tra loro, con servizi e con persone su scala globale. 'L'Internet delle Cose' non è un concetto esclusivamente connesso all'IA. Si tratta infatti di una rete di oggetti fisici – come elettrodomestici, automobili, sensori industriali e dispositivi indossabili – connessi a Internet e capaci di raccogliere e scambiare dati.
L'integrazione tra IoT e IA è spesso indicata con il termine AIoT (*Artificial Intelligence of Things*), che è un paradigma tecnologico che integra intelligenza artificiale (IA) e *Internet of Things* (IoT), applicando algoritmi di elaborazione dei dati e di apprendimento automatico ai flussi informativi generati da dispositivi connessi e dotati di sensori. L'AIoT consente di trasformare grandi quantità di dati raccolti in analisi predittive, automazioni e processi decisionali intelligenti, trovando applicazione in ambiti quali l'industria, l'energia, la sanità e la logistica.

Fonte: Mukhopadhyay S. C., Suryadevara N. K. (2014), Internet of things: Challenges and opportunities, Internet of Things: Challenges and opportunities, pp.1-17

ML (Machine Learning)

L’apprendimento automatico è un insieme di tecniche che permettono ai computer di individuare schemi nei dati e migliorare le prestazioni nel tempo, senza essere programmati in modo dettagliato per ogni singola situazione. Anziché ricevere regole precise da un programmatore, il sistema viene addestrato su molti esempi e impara a riconoscere schemi e regolarità.
Una maggiore quantità di dati pertinenti può contribuire a migliorare i risultati, insieme alla qualità dei dati e alla progettazione del modello. Per esempio, questi sistemi possono essere usati per filtrare e-mail spam, suggerire film o musica, oppure riconoscere la voce.

Fonte: Mitchell T. M. (1997), Machine Learning, McGraw-Hill

Singolarità tecnologica

Un'ipotesi futuristica secondo cui l'intelligenza artificiale potrebbe superare le capacità intellettuali umane, innescando un cambiamento tecnologico e sociale molto rapido e di difficile previsione. È un concetto più filosofico che scientifico, oggetto di ampio dibattito, con posizioni che variano da scenari di possibile realizzazione a valutazioni scettiche sulla sua plausibilità

Fonte: Vinge V. (1993), The Coming Technological Singularity, NASA Conference

Token

Un token è un piccolo pezzo di testo in cui il modello divide una frase per poterla analizzare. A differenza degli umani, che percepiscono frasi e parole come unità naturali, il modello scompone ciò che scriviamo in piccoli pezzi. Questi pezzi possono coincidere con parole intere, ma talvolta sono solo parti di parola o brevi sequenze di caratteri. Per esempio, la parola ‘intelligenza’ potrebbe essere suddivisa in due o tre token, cioè in blocchi distinti che il modello analizza uno alla volta. Oppure, la frase “Ciao, come stai?” può essere segmentata in sette token, poiché alcune parole vengono scomposte in unità sub lessicali (ad esempio radice e desinenza) e i segni di interpunzione sono conteggiati come token autonomi.
In altre parole, il modello non considera la frase come un testo continuo, ma una serie di ‘tessere’ linguistiche, unità minima in cui un modello di IA scompone il testo per poterlo elaborare, che combina per ricostruire il significato complessivo. Sapere cosa sono i token è utile perché i modelli possono elaborare solo un numero limitato di questi pezzi alla volta: il limite non riguarda le pagine o i paragrafi, ma proprio il numero di token che riescono a tenere in memoria in una singola interazione.
Inoltre, è bene evidenziare che i modelli hanno un limite massimo di token che possono gestire in una singola conversazione (la cosiddetta ‘finestra di contesto’). Nei sistemi basati su API, i costi di utilizzo sono spesso calcolati proprio in base al numero di token elaborati. Di conseguenza, la lunghezza dei testi incide non solo sulla capacità del modello di analizzarli, ma anche sui costi e sull’efficienza delle applicazioni di analisi automatica del linguaggio.

Fonte: Jurafsky D., Martin J. H. (2023), Speech and Language Processing, Stanfor University

Context window

La 'finestra di contesto' (*context window*) è la quantità massima di testo, misurata in token, che un modello di intelligenza artificiale può mantenere in memoria in una singola interazione. Può essere paragonata alla memoria di lavoro del modello: tutto ciò che rientra nella finestra viene utilizzato per generare la risposta, ciò che la supera invece, non sarà più considerato nell'elaborazione corrente[13].
Una finestra di contesto più ampia permette al modello l'analisi di documenti più lunghi, nonché il mantenimento di una coerenza in conversazioni articolate e consente di confrontare simultaneamente molte informazioni. Tuttavia, anche nel caso dei modelli con finestre più ampie, se il testo o la conversazione superano il limite disponibile, le parti iniziali possono progressivamente 'uscire' dalla finestra. Pertanto, nell'analisi di testi molto estesi è spesso necessario suddividere i documenti in blocchi (*chunking*) o adottare strategie progressive di sintesi e aggregazione.
Nei modelli più avanzati attualmente disponibili, la finestra di contesto può raggiungere centinaia di migliaia o oltre un milione di token, consentendo l'elaborazione di testi molto estesi in un'unica richiesta, pur restando entro un limite tecnico definito.

Fonte: IBM: https://www.ibm.com/think/topics/context-window
Zhao W. X., Zhou, K., Li J., Tang T., Wang X., Hou Y., Wen J. R. (2023), A survey of Large Language Models, arXiv preprint arXiv:2303.18223, 1(2), 1-124

3.2 Le famiglie dell'intelligenza artificiale: una panoramica

L'intelligenza artificiale consiste in una pluralità di tecnologie tra loro differenti per gli esiti che producono e per la diversa operatività. Tali tecnologie hanno un obiettivo comune che consiste nel consentire alle macchine lo svolgimento di attività tipicamente umane (identificare una immagine, interpretare un linguaggio, effettuare delle scelte in base all'esperienza pregressa etc.).

Le differenti famiglie di IA, IA discriminativa, generativa, predittiva, agentica ed *Embodied* AI (IA incorporata) si distinguono per come sono progettate e per i risultati ai quali conducono. Pertanto, è fondamentale comprenderne le peculiarità per una conoscenza consapevole.

Si consideri, ad esempio, l'IA discriminativa che non genera nuovi contenuti, ma esamina informazioni esistenti e le categorizza. L'IA generativa ha, invece, una funzione creativa dal momento che va oltre la classificazione di dati, producendo testi, immagini, video e musica. Se l'IA discriminativa separa e riconosce, quella generativa crea rendendosi utile alla progettazione e al supporto alla scrittura.

L'IA predittiva è un'ulteriore tipologia di IA focalizzata sulla dimensione temporale. Attraverso l'analisi di dati storici e schemi ricorrenti essa calcola la probabilità che un evento si verifichi (previsioni meteo, tendenze dei mercati finanziari ecc.).

È, invece, una tecnologia progettata per agire in autonomia quella alla base dell'IA agentica che lavora come un vero e proprio assistente operativo. Si tratta di una forma più avanzata di IA che ricevuti obiettivi, pianifica le azioni necessarie al loro perseguimento (es. organizzazione viaggi).

Quando l'intelligenza artificiale viene integrata in un corpo fisico si parla di *Embodied* AI che non opera solo tramite software, ma è inserito in dispositivi come robot o droni, in grado di percepire l'ambiente, muoversi e interagire con gli oggetti (es. magazzini automatizzati, robotica utilizzata in chirurgia).

[13] Un modo semplice per capirla è pensare a un libro. Immaginiamo di poter tenere aperte davanti a noi solo un certo numero di pagine alla volta: se il libro è molto lungo, quando si arriva alle ultime pagine non è possibile più vedere le prime. Non si dimentica tutto per sempre: semplicemente, in quel momento le prime pagine non sono più davanti ai nostri occhi. L'IA funziona allo stesso modo: se il testo è troppo lungo, le parti iniziali possono 'uscire dalla finestra' e non essere più considerate nella risposta.

L’IA, quindi, non rappresenta un sistema unico e indistinto, ma piuttosto un insieme composito di tecnologie con specifiche funzioni e caratteristiche, la cui conoscenza agevola la consapevolezza di potenzialità e limiti nel suo utilizzo.

IA discriminativa

L'IA discriminativa è la forma più consolidata e diffusa: il suo compito è analizzare dati e distinguere tra diverse categorie o classi. È utilizzata, ad esempio, nei sistemi di raccomandazione come quelli di Netflix, ai filtri antispam della posta elettronica e ai sistemi di riconoscimento facciale. In sostanza, questi sistemi analizzano i dati e vengono addestrati a classificarli, distinguendo ad esempio un'e-mail legittima da una di spam, o riconoscendo un gatto in una fotografia.

Fonte: Deng L., O'Shaughnessy D. (2015), Deep Discriminative and Generative Models, In Handbook of Speech Processing, Microsoft Research
Sarker I. H. *et al.* (2022), AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems, SN Computer Science, 3(158)

IA generativa

La branca dell'intelligenza artificiale in grado di creare contenuti originali – testi, immagini, musica, video e codice – che prima non esistevano. A differenza dell'IA tradizionale, che classifica o analizza dati esistenti, l'IA generativa produce qualcosa di nuovo a partire da un input fornito dall’utente, comunemente definito prompt. Il prompt consiste in una richiesta o istruzione, espressa in linguaggio naturale o in altra forma strutturata, che guida il modello nella generazione dell’output.
Diversamente dai sistemi di IA orientati alla classificazione o alla previsione, l’IA generativa produce contenuti originali coerenti con le regolarità apprese durante l’addestramento. In termini generali, il funzionamento può essere descritto così: a una richiesta dell’utente corrisponde la generazione di un output coerente con i dati di addestramento.
Esempi di applicazioni di IA generativa includono modelli linguistici come ChatGPT e Claude per la generazione di testo, nonché sistemi come DALL·E e Midjourney per la creazione di immagini.

Fonte: Gartner Glossary: Generative AI, cfr. Istat (2024), Digitalizzazione, interoperabilità e intelligenza artificiale, Diritto delle nuove tecnologie E-book
Feuerriegel S. *et al.* (2024), Generative AI, Business & Information Systems Engineering, 66(1), pp.111-126

IA predittiva

L'IA predittiva utilizza dati storici e modelli statistici per stimare la probabilità di eventi futuri. È impiegata in ambiti come le previsioni meteorologiche, l'analisi dei mercati finanziari, la manutenzione preventiva degli impianti industriali e la medicina personalizzata. A differenza dell'IA generativa, non crea contenuti nuovi, ma formula previsioni basate su tendenze e correlazioni individuate nei dati.

Fonte: Hasan U., Shreevamshi N. (2025), AI-Powered Predictive Analytics for Financial Forecasting and Strategic Insight, International Journal of Research and Innovation in Applied Science, X. 532-555. 10.51584/IJRIAS.2025.10060039

IA agentica

L'IA agentica rappresenta la frontiera più recente dell'evoluzione tecnologica. Mentre i chatbot tradizionali si limitano a rispondere alle domande dell'utente, un agente di IA può eseguire in modo automatizzato una sequenza di azioni per soddisfare un obiettivo definito dall'utente: può navigare su internet, consultare database, eseguire operazioni e coordinare altri agenti in sistemi multi-agente complessi. L'utente definisce il traguardo e l'agente seleziona automaticamente le azioni ritenute più appropriate in base ai criteri di progettazione e ai dati disponibili, con una capacità di pianificazione e adattamento che si avvicina a quella di un collaboratore umano. Per una definizione tecnica del concetto di agente di IA si veda la voce 'AI Agent'.

Fonte: Nisa U., Shirazi M., Saip M.A., Pozi M.S.M., (2025), Agentic AI: The Age of Reasoning, A review. Journal of Automation and Intelligence, 10.1016/j.jai.2025.08.003

Embodied AI (IA incorporata)

Infine, l'IA incorporata (*Embodied* AI) è quella integrata in sistemi fisici come robot, droni e dispositivi autonomi. Combina percezione sensoriale, capacità decisionale e azione fisica, ed è alla base della robotica autonoma, dei veicoli a guida automatica e dei sistemi industriali intelligenti.

Fonte: Lisondra M., Benhabib B., Nejat G. (2025), Embodied AI with Foundation Models for Mobile Service Robots: A Systematic Review, arXiv preprint arXiv:2505.20503
Feng T., Wang X., Jiang Y. G., Zhu W. (2025), Embodied AI: From LLMs to World Models, arXiv preprint arXiv:2509.20021

Queste famiglie non sono compartimenti stagni: nella pratica, i sistemi più avanzati combinano elementi di diverse categorie. Un robot domestico, ad esempio, può utilizzare l'IA discriminativa per riconoscere gli oggetti, l'IA generativa per comunicare in linguaggio naturale e l'IA agentica per pianificare le proprie azioni.

4. Modelli generativi e principali attori

4.1 IA generativa e modelli linguistici (LLM)

Questa sezione rappresenta un punto chiave della rivoluzione tecnologica che, a partire dal 2022, ha portato l'intelligenza artificiale al centro del dibattito pubblico globale: l'IA generativa e i grandi modelli linguistici (*Large Language Models*, LLM). La prima categoria ha fornito le basi concettuali, in questa, invece, si progredisce entrando più nello specifico delle tecnologie e degli strumenti che stanno trasformando il modo in cui lavoriamo, comunichiamo e creiamo contenuti.

Il paragrafo si apre con le definizioni dei concetti chiave che caratterizzano questa famiglia di tecnologie: cosa si intende per IA generativa (GenAI), come funzionano i modelli linguistici di grandi dimensioni e quale ruolo giocano elementi tecnici come la finestra di contesto, i prompt e le tecniche di prompt engineering nell'interazione tra l'essere umano e la macchina.

In questa parte del documento si comprenderà perché la qualità di una richiesta formulata a un modello influenzi profondamente la qualità della risposta ottenuta, e si apprenderanno tecniche come la *Retrieval-Augmented Generation* (RAG), che migliora l'accuratezza delle risposte combinando generazione e recupero di informazioni da fonti esterne.

Ampio spazio è dedicato ai protagonisti tecnologici (senza alcuna pretesa di esaustività): da ChatGPT di OpenAI, il prodotto che ha segnato l'ingresso dell'IA generativa nella vita quotidiana di milioni di persone, a Claude di Anthropic, progettato con un'attenzione particolare alla sicurezza e all'allineamento etico. Si illustra poi un panorama di alcuni degli attori principali che competono nel mercato globale dell'IA generativa, ciascuno con filosofie, architetture e modelli di business differenti e vengono descritte Gemini (Google), Copilot (Microsoft), DeepSeek, Grok, Llama (Meta), Mistral e Perplexity.

Il paragrafo è rilevante per chi ha l'obiettivo di comprendere non solo cosa sono questi strumenti, ma anche come utilizzarli in modo consapevole e quali sono le differenze fondamentali tra i vari modelli disponibili sul mercato. Al suo interno, si troveranno le chiavi di lettura per orientarsi in un ecosistema tecnologico in rapidissima evoluzione, dove l'elaborazione del linguaggio naturale (NLP) rappresenta un aspetto chiave.

S'intende segnalare, ai fini di una comprensione più esaustiva delle caratteristiche dei modelli linguistici, che questi ultimi, essendo basati sull'intelligenza artificiale, non 'pensano' nel senso umano del termine. Funzionano perché sono addestrati su una grandissima quantità di testi e hanno imparato a riconoscere regolarità e schemi nel linguaggio. In altri termini si può affermare che il modello impara schemi relativi a quali parole tendono a comparire insieme e in quale ordine, così da poter prevedere e generare la parola successiva in modo coerente. Pertanto, se anche i modelli linguistici basati su intelligenza artificiale possano essere in grado di generare risposte con un tono percepito come empatico, ciò avviene perché sono addestrati su grandi quantità di testi in cui compaiono espressioni di rassicurazione, comprensione e supporto, e apprendono le regolarità linguistiche associate a tali modalità comunicative. Non provano emozioni né possiedono stati mentali, ma riproducono schemi ricorrenti del linguaggio umano.

La percezione di empatia che l'utente pertanto potrebbe avere deriva dall'interazione tra modelli statistici del linguaggio e meccanismi cognitivi umani e non da un'esperienza emotiva reale da parte del modello (cosiddetto antropomorfismo).

La letteratura mostra che esiste una relazione tra il livello di antropomorfismo percepito (cioè quanto un'IA sembra "umana"), la presenza sociale (la sensazione che l'agente sia davvero presente e interagisca con noi) e l'esperienza dell'utente. Secondo Williams (2025), questa relazione non ha sempre la stessa intensità: è più forte nel caso di robot di servizio e avatar, mentre risulta solo moderatamente significativa per chatbot e assistenti vocali.

Questa sezione è ordinata in diverse gradazioni che possono sostenere un percorso graduale di conoscenza degli aspetti principali dei modelli linguistici: la prima riguarda i concetti chiave dei LLM, la seconda riguarda l'infrastruttura tecnologica, la terza descrive alcuni esempi di interfacce e prodotti commerciali diffusisi negli anni più recenti (senza alcuna pretesa di esaustività ed esclusivamente a titolo esemplificativo).

4.1.1 Concetti chiave dei LLM

GenAI (IA Generativa)

Vedi capitolo 3.2

GPT (Generative Pre-trained Transformer)

I GPT (Transformer Generativo Pre-addestrato) sono una famiglia di modelli linguistici di grandi dimensioni (LLM) sviluppati da OpenAI e basati su un'architettura Transformer di tipo *decoder-only*. Sul piano lessicale, il termine è talvolta utilizzato in senso più ampio per riferirsi, in generale, a modelli generativi fondati su *architetture Transformer*.
I modelli GPT trovano applicazione nella generazione e analisi del linguaggio naturale, nella sintesi di contenuti, nella traduzione, nella programmazione assistita e in sistemi multimodali in grado di elaborare testo, immagini e altri tipi di input.
Il loro funzionamento si fonda su una fase iniziale di *pre-training* (pre-addestramento) su enormi quantità di dati testuali, finalizzata all'apprendimento delle regolarità statistiche del linguaggio. Successivamente, il modello può essere ulteriormente adattato tramite un addestramento aggiuntivo (*fine-tuning*) oppure utilizzato direttamente per generare risposte coerenti a nuove richieste.
L'architettura Transformer, su cui i GPT si basano, è stata introdotta da Vaswani *et al*. (2017) nel lavoro *Attention Is All You Need* e si fonda sul meccanismo di *self-attention* per l'elaborazione parallela delle sequenze. A partire da questa architettura, OpenAI ha sviluppato la serie GPT (dal GPT-1 nel 2018 fino alle versioni successive), potenziandone progressivamente dimensioni e capacità generative.

Fonte: Vaswani A*. et al.* (2017), Attention is All You Need; NeurIPS; U.S. Patent & Trademark Office

Brown T. B*. et al*. (2020), Language Models are Few-Shot Learners, NeurIPS; U.S. Patent & Trademark Office, cfr. IBM Think: https://www.ibm.com/it-it/think/topics/gpt

LLM (Large Language Model)

Gli LLM sono modelli statistici progettati per stimare la parola successiva in una sequenza di testo. Vengono addestrati su grandi quantità di testo per riconoscere schemi ricorrenti e usarli per generare nuove frasi. Si tratta di programmi di intelligenza artificiale addestrati su enormi quantità di testo (libri, siti web, articoli) per imparare a comprendere e generare il linguaggio umano. In generale, modelli con un numero maggiore di parametri possono mostrare prestazioni linguistiche più avanzate, anche se i risultati dipendono anche dai dati e dal modo in cui sono stati addestrati. Costituiscono la tecnologia alla base di assistenti oggi accessibili al pubblico, come ChatGPT, Claude e Gemini, e rappresentano il risultato di decenni di progressi nell'elaborazione del linguaggio naturale (NLP) e nella ricerca sull'apprendimento automatico, in particolare a partire dal 2010.

Fonte: https://www.ibm.com/it-it/topics/large-language-models

NLP (Natural Language Processing)

L'elaborazione del linguaggio naturale (*Natural Language Processing*, NLP) è il settore dell'intelligenza artificiale che sviluppa metodi e modelli per l'analisi, la rappresentazione e la generazione del linguaggio umano in forma testuale o vocale. Comprende attività quali il riconoscimento del parlato, la traduzione automatica, la comprensione del testo, la sintesi e il riassunto di documenti, nonché l'analisi del sentimento espresso in un messaggio.
Le applicazioni di NLP sono oggi diffuse in numerosi strumenti digitali, tra cui assistenti vocali, sistemi di traduzione automatica e piattaforme di analisi testuale.

Fonte: Jurafsky D., Martin J. H. (2023), Speech and Language Processing, 3rd edition, Stanford University

Prompt

Il prompt è un input che può assumere la forma di una domanda, di un'istruzione, di un testo contestuale, di un esempio o di una combinazione di questi elementi.
È il punto di partenza di ogni interazione con strumenti come ChatGPT, Claude o Gemini. La qualità del prompt può incidere sulla qualità dell'output: richieste più chiare e specifiche tendono a produrre risultati più pertinenti.

Fonte: https://www.treccani.it/vocabolario/prompt/

Prompt Engineering

Insieme di tecniche e strategie volte a formulare le richieste (prompt) a un modello di IA nel modo più efficace possibile per ottenere risposte ottimali. Include strategie come fornire esempi, specificare il formato desiderato, assegnare un ruolo al modello e suddividere compiti complessi in passaggi più semplici.
Il *prompt engineering* può essere considerato una competenza emergente di progettazione dell'interazione uomo-IA, che combina capacità linguistiche, comprensione del funzionamento dei modelli e definizione strategica degli obiettivi.
È opportuno evidenziare che il prompt engineering in senso professionale implica un livello avanzato di progettazione dell'interazione con i modelli linguistici e consiste nella progettazione strategica e sistematica dell'interazione con modelli di IA, finalizzata all'integrazione efficace di tali sistemi in processi produttivi o decisionali complessi.

Fonte: Bolognesi F. (2024), Intelligenza artificiale, Istruzioni per l'uso, EPC Editore

RAG (Retrieval-Augmented Generation)

La 'Generazione Aumentata dal Recupero di Informazioni' è un'architettura che combina modelli generativi con un sistema di recupero di informazioni da fonti esterne. Sono integrati quindi due passaggi: il modello combina le informazioni apprese durante l'addestramento con la ricerca di informazioni aggiornate e pertinenti in un archivio di documenti, poi grazie a queste informazioni la risposta è più accurata e fondata. È come se l'IA consultasse un'enciclopedia prima di rispondere, piuttosto che affidarsi solo alla propria 'memoria'. Questo consente di migliorare l'accuratezza, l'aggiornamento e la tracciabilità delle risposte, riducendo il rischio di informazioni imprecise o non fondate.
Le fonti possono essere interne all'organizzazione o, se configurato, provenire da basi informative esterne; tuttavia, in molti contesti aziendali e istituzionali l'architettura RAG è utilizzata principalmente per interrogare documentazione proprietaria o validata.
In ambito normativo, le architetture RAG sono utilizzate per interrogare basi documentali ufficiali (es. testi di legge, linee guida, sentenze), consentendo la generazione di risposte fondate su fonti citabili e aggiornate. In ambito accademico, strumenti analoghi sono impiegati per la revisione della letteratura, combinando il recupero di articoli scientifici rilevanti con la sintesi automatizzata dei contenuti.

Fonte: Lewis P. *et al.* (2020), Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, Advances in Neural Information Processing Systems, 33

CoT (Chain of Thought)

La CoT (*Chain of Thought*, catena di pensiero) è una strategia di *prompting* che incoraggia il modello a generare passaggi intermedi espliciti prima di fornire una risposta. È particolarmente utile per compiti che richiedono più passaggi logici o di calcolo. Formulazioni del tipo 'spiega i passaggi' o 'mostra il ragionamento' possono migliorare la qualità delle risposte, rendendo più trasparente il percorso che conduce all'esito finale.

Fonte: Ton J. F. *et al.* (2024), Understanding Chain-of-Thought in LLMs Through Information Theory, arXiv preprint arXiv:2411.11984

4.1.2 Famiglie di modelli (infrastruttura tecnologica)

GPT (Generative Pre-trained Transformer)

Vedi GTP capitolo 4.1.1.

Llama

Famiglia di modelli linguistici open-source sviluppati da Meta (la società di Facebook). Essendo rilasciati con licenze che ne consentono l'uso e l'adattamento in diversi contesti, i modelli Llama possono essere scaricati e personalizzati da sviluppatori e ricercatori di tutto il mondo, favorendo una più ampia diffusione della ricerca e dello sviluppo nel campo dei modelli linguistici. A differenza di servizi proprietari, i modelli Llama possono essere integrati in infrastrutture locali e personalizzati per specifiche esigenze applicative, favorendo sperimentazione, ricerca indipendente e sviluppo di soluzioni aziendali su misura.

Fonte: https://ai.meta.com/blog/large-language-model-llama-meta-ai/

Mistral

Famiglia di modelli linguistici sviluppata dalla società francese Mistral AI. I modelli Mistral sono progettati con particolare attenzione all'efficienza computazionale e all'ottimizzazione delle prestazioni in rapporto al numero di parametri, con l'obiettivo di rendere i *Large Language Models* più accessibili e sostenibili dal punto di vista infrastrutturale. Si collocano nel panorama europeo dello sviluppo di modelli linguistici di grandi dimensioni.

Fonte: https://www.mistral.ai/

Gemini

Gemini è una famiglia di modelli di intelligenza artificiale sviluppata da Google, basata su architetture di tipo Transformer e progettata per applicazioni multimodali, ossia in grado di elaborare e generare contenuti testuali, visivi e, in alcune versioni, audiovisivi. I modelli Gemini sono integrati nell'ecosistema dei servizi Google (Gmail e Google Workspace).

Fonte: https://gemini.google.com/

Grok

Famiglia di modelli linguistici di grandi dimensioni sviluppata da xAI, l'azienda di Elon Musk. È integrato nella piattaforma X (ex Twitter) e si caratterizza per uno stile di comunicazione diretto e informale.
Come altri *Large Language Models*, Grok si basa su architetture neurali di tipo Transformer e viene impiegato per la generazione e l'elaborazione di testo in linguaggio naturale.

Fonte: Grok 3: https://x.ai/

DeepSeek

Azienda cinese che ha attirato l'attenzione mondiale nel 2025 per i suoi modelli linguistici avanzati e competitivi, sviluppati con costi significativamente inferiori rispetto ai concorrenti occidentali. I suoi modelli, come DeepSeek-V3 e R1, sono utilizzati per la generazione di contenuti e la ricerca. La loro comparsa ha sollevato importanti discussioni sulla competizione globale nel campo dell'IA.

Fonte: https://www.deepseek.com/

4.1.3 Interfacce ed esempi di prodotti commerciali

ChatGPT

ChatGPT è un sistema conversazionale basato su modelli linguistici della famiglia GPT, sviluppati da OpenAI. È progettato per interagire in linguaggio naturale con l'utente, generando risposte testuali coerenti a partire da input forniti sotto forma di prompt.
Utilizza modelli di intelligenza artificiale generativa addestrati su grandi quantità di dati testuali e può svolgere compiti quali rispondere a domande, redigere testi, tradurre, sintetizzare contenuti e generare codice.
A partire dal 2022, ChatGPT ha contribuito in modo significativo alla diffusione pubblica dell'IA generativa.

Fonte: https://openai.com/blog/chatgpt/

Claude

Assistente virtuale basato sull'intelligenza artificiale sviluppato da Anthropic, progettato per interagire con gli utenti attraverso conversazioni in linguaggio naturale.
È presentato dall'azienda come orientato alla sicurezza e all'allineamento con principi di utilizzo responsabile. Disponibile in diverse versioni (Claude Sonnet, Claude Opus), è utilizzato sia da utenti privati che da aziende.

Fonte: https://www.anthropic.com/index/claude

Copilot (Microsoft Copilot)

Assistente di intelligenza artificiale generativa integrato nei prodotti Microsoft come Word, Excel, PowerPoint, Outlook e il motore di ricerca Bing. Aiuta gli utenti a scrivere testi, creare presentazioni, analizzare dati e automatizzare attività quotidiane attraverso comandi in linguaggio naturale. In pratica, Copilot è come un collega digitale che lavora all'interno dei programmi Microsoft.

Fonte: Microsoft Copilot: https://copilot.microsoft.com/onboarding

Perplexity

Motore di ricerca conversazionale che utilizza l'intelligenza artificiale generativa per rispondere alle domande degli utenti in modo articolato, accompagnando le risposte con riferimenti alle fonti. A differenza di un motore di ricerca tradizionale che mostra una lista di link, Perplexity fornisce direttamente una risposta ragionata e referenziata.

Fonte: https://www.perplexity.ai/

4.2 Strumenti creativi e piattaforme per la generazione di contenuti multimediali

L'intelligenza artificiale generativa non si limita alla creazione di testo: questa sezione accompagna nella conoscenza degli strumenti e delle piattaforme che hanno esteso le capacità creative dell'IA al mondo delle immagini, dei video, della musica e della voce. Si tratta di un ecosistema in espansione che sta ridefinendo i confini della produzione multimediale, rendendo accessibili a tutti attività che fino a pochi anni fa richiedevano competenze specialistiche e risorse significative.

In questa sezione sono descritti strumenti di generazione audiovisiva basati su modelli generativi, articolati in:

- sistemi per la creazione di immagini a partire da input testuale o visivo;
- strumenti per la generazione e trasformazione di contenuti video;
- soluzioni per la sintesi vocale, la clonazione della voce e la generazione musicale;
- piattaforme per la produzione di contenuti con avatar digitali;
- strumenti per l'analisi e l'interazione con documenti su corpus chiuso.

Gli esempi citati hanno finalità esclusivamente illustrativa e non intendono costituire una rassegna esaustiva del mercato; sono selezionati per rappresentare in modo concreto le diverse categorie tecnologiche e le loro principali funzioni.

Questa sezione potrebbe essere particolarmente utile per professionisti della comunicazione, formatori, creativi e chiunque desideri comprendere come l'IA stia trasformando la produzione di contenuti multimediali, quali strumenti siano disponibili e quali opportunità e sfide comporti questa nuova frontiera della creatività assistita dall'intelligenza artificiale.

4.2.1 Generazione delle immagini

Questa categoria comprende strumenti basati su modelli generativi capaci di generare immagini a partire da descrizioni testuali (prompt) o da immagini di input. L'utente fornisce un'istruzione in linguaggio naturale e il sistema genera un contenuto visivo sulla base delle indicazioni fornite.

Nel linguaggio comune tali strumenti sono indicati anche come AI Art Generators.

Esempi includono DALL·E, Midjourney, Stable Diffusion, Adobe Firefly e Nano Banana.

4.2.2 Generazione e trasformazione video

Questa categoria comprende sistemi basati su modelli generativi in grado di generare o modificare contenuti video a partire da descrizioni testuali, immagini o sequenze audiovisive preesistenti.

Tali strumenti possono operare sia in modalità text-to-video, producendo nuove sequenze coerenti con un prompt linguistico, sia in modalità di trasformazione e editing di video esistenti.

Esempi illustrativi includono Sora, Runway e Pika.

4.2.3 Generazione vocale e sintesi del parlato

Strumenti che producono voce sintetica o clonano voci esistenti.

Questa categoria comprende strumenti basati su modelli di sintesi vocale e di voice cloning con cui è possibile generare voce sintetica a partire da testo scritto o di replicare caratteristiche vocali specifiche mediante tecniche di apprendimento automatico.

Esempi illustrativi includono ElevenLabs e PlayHT.

4.2.4 Generazione musicale e sonora

Questa categoria comprende strumenti basati su modelli generativi che producono musica, effetti sonori o composizioni complete a partire da input testuali e parametri stilistici. Tali sistemi si basano su modelli neurali addestrati su grandi corpus musicali per generare sequenze audio coerenti sotto il profilo armonico, ritmico e timbrico.

Esempi illustrativi includono Suno e Udio.

4.2.5 Avatar digitali e video sintetici con personaggi

Questa categoria comprende sistemi che combinano modelli linguistici, sintesi vocale e rappresentazione visiva di personaggi virtuali per la produzione di video con avatar digitali parlanti. Tali strumenti integrano generazione testuale, sintesi del parlato e animazione facciale automatizzata.

Esempi illustrativi includono Synthesia e HeyGen.

4.2.6 Strumenti multimodali integrati

Questa categoria comprende piattaforme basate su modelli multimodali in grado di elaborare e generare contenuti su più modalità (testo, immagini, audio, video) all'interno di un'unica architettura. Tali sistemi consentono interazioni cross-modali, ossia la trasformazione o integrazione di informazioni provenienti da formati differenti.

Esempi illustrativi includono Gemini, GPT-4o e alcune piattaforme che integrano funzionalità multimodali in ambienti di produzione creativa.

4.2.7 Strumenti di co-creazione e integrazione nei workflow

Questa categoria comprende applicazioni di intelligenza artificiale progettate per supportare il processo creativo e decisionale attraverso l'integrazione diretta nei flussi di lavoro professionali. A differenza dei sistemi puramente generativi, tali strumenti non operano esclusivamente nella produzione automatizzata di contenuti, ma affiancano l'utente in attività di progettazione, scrittura, revisione e organizzazione dell'informazione.

Si tratta di soluzioni che incorporano modelli generativi all'interno di ambienti software già utilizzati nei contesti lavorativi, favorendo forme di collaborazione uomo-macchina nei flussi di lavoro professionali e nei contesti produttivi.

Esempi illustrativi includono Canva e Microsoft Copilot.

4.2.8 Strumenti di analisi e interazione documentale

Questa categoria comprende sistemi progettati per l'analisi e l'interazione con corpus documentali chiusi, ossia insiemi di testi forniti o caricati dall'utente. A differenza dei modelli generativi general purpose, tali strumenti operano su basi informative delimitate, orientando la generazione delle risposte ai contenuti effettivamente presenti nei documenti.

Rientrano in questa tipologia:

Sistemi di summarization avanzata, che consentono la sintesi di documenti complessi attraverso tecniche di tipo estrattivo (selezione delle parti rilevanti del testo originale) o astrattivo (riformulazione sintetica dei contenuti).

Piattaforme di question answering su documenti, che permettono di interrogare un corpus tramite domande in linguaggio naturale. Tali sistemi combinano tecniche di retrieval semantico con modelli linguistici generativi, spesso secondo architetture di tipo RAG (Retrieval-Augmented Generation), al fine di individuare le porzioni informative pertinenti e formulare risposte coerenti.

Esempi illustrativi includono strumenti come NotebookLM.

4.3 Agenti di intelligenza artificiale e sistemi autonomi

Con questo paragrafo si entra nella parte più avanzata dello scenario attuale inerente all'intelligenza artificiale: quello degli agenti IA e dei sistemi capaci di agire in modo autonomo nel mondo digitale e fisico. Finora, di fatto, sono stati illustrati strumenti che rispondono alle richieste dell'utente, qui il focus si sposta su sistemi che possono pianificare, decidere e agire con gradi crescenti di indipendenza.

Il concetto centrale è quello relativo all'AI Agent: un programma di intelligenza artificiale che va ben oltre il semplice chatbot, essendo in grado di navigare sul web, utilizzare strumenti software, concatenare azioni e prendere decisioni operative per raggiungere obiettivi definiti dall'utente, con poca o nessuna supervisione durante l'esecuzione. A questo si affiancano le voci dedicate a due approcci emergenti nello sviluppo software: il *vibe coding* è basato su un'interazione conversazionale e intuitiva con l'IA, e *l'agentic coding*, in cui l'agente IA pianifica e realizza autonomamente soluzioni software complete.

La sezione esplora inoltre la robotica autonoma, che rappresenta una sorta di combinazione tra intelligenza artificiale e macchine capaci di percepire l'ambiente, prendere decisioni e compiere azioni concrete: dalle automobili a guida autonoma ai robot industriali, dai droni alle applicazioni in ambito chirurgico. Accanto a tali concetti, saranno spiegate tipologie di tecnologie più consolidate che automatizzano attività ripetitive e basate su regole all'interno dei processi aziendali, e gli assistenti virtuali come Siri, Alexa e Google Assistant, che rappresentano la forma più familiare di agente IA nella vita quotidiana.

Completa il paragrafo la funzionalità con cui Google ha integrato l'IA direttamente nel motore di ricerca, trasformando l'esperienza di ricerca da una lista di link a una risposta ragionata. In questa categoria sarà possibile rintracciare gli elementi per comprendere la direzione in cui si sta muovendo l'IA: verso sistemi sempre più autonomi e capaci di agire nel mondo, con tutte le opportunità e le responsabilità che questo comporta per i produttori e per i fruitori.

Agentic Coding

L'*agentic coding* indica l'impiego di agenti di intelligenza artificiale nello sviluppo software. In questo approccio, il programmatore definisce obiettivi e vincoli del progetto, mentre il sistema può generare, testare e modificare automaticamente il codice sulla base delle specifiche fornite. L'intervento umano rimane centrale nella definizione dei requisiti, nella supervisione e nella validazione finale, mentre le attività operative di scrittura e revisione del codice possono essere in parte automatizzate. *L'agentic coding* rappresenta una recente evoluzione nelle pratiche di automazione dello sviluppo software.

Fonte: Sapkota R., Roumeliotis, K. I., Karkee, M. (2025), Vibe Coding vs. Agentic Coding: Fundamentals and Practical Implications of Agentic AI, arXiv preprint arXiv:2505.19443

AI Agent

L'agente AI è un sistema di intelligenza artificiale progettato per operare in funzione di uno o più obiettivi definiti attraverso una sequenza di azioni automatizzate o semi-automatizzate. A differenza di un semplice chatbot che si limita a generare risposte testuali, un agente IA può organizzare automaticamente una sequenza di operazioni, utilizzare strumenti esterni (software, API, browser) e selezionare passaggi intermedi, ad esempio per analizzare dati, automatizzare processi o interagire con altre applicazioni. Il livello di supervisione umana varia in base al contesto applicativo. Gli AI agent sono spesso basati su LLM integrati con memoria, strumenti esterni e meccanismi di gestione delle azioni.

Fonte: IBM: https://www.ibm.com/think/topics/ai-agents#What+are+ai+agents?

AI Overviews

La funzionalità integrata da Google nel suo motore di ricerca che utilizza l'intelligenza artificiale per presentare una panoramica sintetica e generata automaticamente sull'argomento cercato, accompagnata da link alle fonti originali per approfondire. Piuttosto che ottenere come risultato una lista di link, il motore di ricerca spiega direttamente la risposta rimandando alle fonti consultate nella composizione della risposta.
È una evoluzione dei tradizionali motori di ricerca e utilizza modelli di IA generativa.

Fonte: Google blog: https://blog.google/feed/were-bringing-the-helpfulness-of-ai-overviews-to-more-countries-in-europe/

Robotica autonoma

Il campo che unisce intelligenza artificiale e robotica per creare macchine capaci di percepire l'ambiente circostante, prendere decisioni e compiere azioni fisiche, senza o con livelli variabili di intervento umano. Tali sistemi combinano sensori, algoritmi di controllo e modelli di apprendimento automatico per adattarsi a contesti dinamici senza intervento umano diretto continuo. Dalle auto a guida autonoma ai droni per le consegne, dai robot industriali ai robot chirurgici, è una delle applicazioni più tangibili dell'IA nel mondo fisico

Fonte: Siciliano B., Khatib O. (2016), Springer Handbook of Robotics. 2nd edition

RPA (Robotics Process Automation)

È una tecnologia che consente di automatizzare attività digitali ripetitive e strutturate mediante l'utilizzo di 'software robot' progettati per eseguire attività ripetitive e basate su regole che normalmente vengono svolte da persone, come l'inserimento di dati, la compilazione di moduli o la gestione di fatture.
I sistemi RPA eseguono automaticamente sequenze di azioni su applicazioni e interfacce digitali seguendo flussi predefiniti.

Fonte: IEEE Standards Association: RPA

Vibe Coding

Un approccio allo sviluppo software in cui il programmatore interagisce con l'IA in modo conversazionale e intuitivo, descrivendo in linguaggio naturale (con semplici parole) le funzionalità desiderate. Il sistema genera codice sulla base delle istruzioni fornite, mentre l'utente interviene per verificarlo, modificarlo e perfezionarlo in iterazioni successive.
A differenza dell'*agentic coding*, in cui l'agente pianifica ed esegue autonomamente sequenze operative, il *vibe coding* permette una interazione continua tra umano e IA, privilegiando sperimentazione, velocità di prototipazione e inventiva rispetto alla scrittura manuale tradizionale del codice.

Fonte: Sapkota R., Roumeliotis K. I., Karkee, M. (2025), Vibe Coding vs. Agentic Coding: Fundamentals and Practical Implications of Agentic AI, arXiv preprint arXiv:2505.19443

Virtual Assistant

Sistema conversazionale basato su tecnologie di intelligenza artificiale in grado di elaborare input vocali o testuali e di rispondere tramite testo o voce sintetizzata. È integrato in dispositivi come smartphone, smart speaker e altre piattaforme digitali e consente l'esecuzione di comandi per ottenere informazioni, controllare dispositivi connessi, gestire contenuti multimediali e svolgere attività quotidiane (e-mail, calendari, promemoria).
Tra gli assistenti virtuali più diffusi, in particolare nella versione vocale, si annoverano Siri (Apple), Alexa (Amazon), Cortana (Microsoft) e Google Assistant.

Fonte: Hoy M. B. (2018), Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants, Medical Reference Services Quarterly, 37(1), pp.81-88

4.4 Aziende e protagonisti dell'intelligenza artificiale

Quest'area della tassonomia offre una mappa delle aziende e delle figure chiave che hanno plasmato il campo dell'intelligenza artificiale, dalle origini storiche fino all'attuale competizione globale tra laboratori di ricerca e colossi tecnologici. Conoscere questi protagonisti è essenziale per comprendere non solo le origini dell'IA, ma anche le dinamiche economiche, strategiche e ideologiche che ne hanno guidato finora lo sviluppo.

Sul versante storico, si citeranno i padri fondatori della disciplina: Alan Turing, il matematico britannico che pose le basi teoriche dell'intelligenza delle macchine, e John McCarthy, che nel 1956 coniò il termine 'Intelligenza Artificiale' al Dartmouth College, luogo di nascita ufficiale della disciplina. Si arriva poi ai protagonisti della rinascita moderna dell'IA: Geoffrey Hinton, studioso cardine del 'Deep Learning' insignito del Premio Nobel per la Fisica nel 2024, e Yann LeCun, padre delle reti neurali convoluzionali e Premio Turing 2018.

Sul fronte industriale, la sezione presenta le aziende che oggi guidano la corsa all'IA:

- OpenAI, creatrice di ChatGPT e tra le realtà più influenti del settore;
- Anthropic, focalizzata sulla sicurezza e l'allineamento etico;
- DeepMind di Google, celebre per risultati pionieristici come AlphaGo e AlphaFold;
- Meta, che con la famiglia Llama ha scommesso sull'open source come strategia di sviluppo;
- NVIDIA, il cui dominio nel mercato delle GPU l'ha resa l'infrastruttura hardware indispensabile dell'intera industria dell'IA.

Completa il quadro Hugging Face, la piattaforma open source che ha democratizzato l'accesso a modelli e risorse, diventando il punto di riferimento per sviluppatori e ricercatori di tutto il mondo.

Questa parte della tassonomia permette di comprendere l'ecosistema dell'intelligenza artificiale nella sua dimensione umana e organizzativa, riconoscendo i contributi individuali e collettivi che hanno reso possibile la rivoluzione tecnologica in corso e le tensioni che ne caratterizzano l'evoluzione: tra apertura e proprietà intellettuale, tra innovazione e sicurezza, tra profitto e bene comune.

AI Winter

'L'Inverno dell'IA' è un periodo storico in cui l'interesse e i finanziamenti per la ricerca sull'intelligenza artificiale sono calati drasticamente, a causa di aspettative troppo alte e risultati deludenti. Si ricordano due periodi molto negativi: il primo a metà degli anni '70, il secondo tra la fine degli anni '80 e l'inizio degli '90. Questi cicli evidenziano che il progresso dell'IA non è stato lineare, ma è stato caratterizzato da fasi di maggior investimento e periodi di riduzione di risorse nell'interesse scientifico.

Fonte: Crevier D. (1993), AI: The Tumultuous History of the Search for Artificial Intelligence, Basic Books

Anthropic

Azienda americana di ricerca sull'intelligenza artificiale, fondata da ex dipendenti di OpenAI, focalizzata sullo sviluppo di modelli linguistici avanzati con un orientamento dichiarato alla sicurezza e all'allineamento ai valori umani e all'utilizzo responsabile. È sviluppatrice dell'assistente Claude.

Fonte: https://www.anthropic.com/

Dartmouth College

È generalmente considerato il luogo di avvio formale dell'intelligenza artificiale come disciplina accademica. Nel 1956 vi si tenne la *Dartmouth Summer Research Project on Artificial Intelligence*, la conferenza in cui ricercatori come McCarthy, Minsky e Shannon proposero il termine '*Artificial Intelligence*' e ne delinearono gli obiettivi di ricerca.

Fonte: McCarthy J., Minsky, M. L. Rochester, N., Shannon C. E. (2006), A proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955, AI magazine, 27(4), pp.12-12

DeepMind

Laboratorio di ricerca sull'intelligenza artificiale di proprietà di Google (Alphabet), nato nel 2010 a Londra. È nota per progetti come AlphaGo (il programma che ha battuto il campione mondiale di Go nel 2016), AlphaFold (che ha rivoluzionato la previsione della struttura delle proteine) e numerosi altri contributi all'avanguardia della ricerca sull'IA.

Fonte: https://deepmind.google/

Hinton, Geoffrey

Informatico britannico-canadese spesso indicato come uno dei principali pionieri del deep learning. Il suo lavoro sulla *backpropagation* (retro-propagazione dell'errore) ha reso possibile l'addestramento efficace delle reti neurali profonde, gettando le basi per la rivoluzione dell'IA moderna. Nel 2024 ha ricevuto il Premio Nobel per la Fisica per i suoi contributi. Oggi è una voce autorevole sui rischi dell'IA avanzata.

Fonte: https://www.treccani.it/enciclopedia/eol-hinton-geoffrey-everest/

Hugging Face

La più grande piattaforma e comunità open-source per l'intelligenza artificiale. Ospita migliaia di modelli, dataset e strumenti, molti dei quali accessibili secondo licenze open-source o con modalità di utilizzo pubbliche, permettendo a sviluppatori e ricercatori di tutto il mondo di condividere, scoprire e utilizzare risorse per i loro progetti di IA. È considerata il 'GitHub dell'IA'.

Fonte: https://huggingface.co/

LeCun, Yann

Informatico franco-americano, responsabile della ricerca IA presso Meta. È tra i principali contributori allo sviluppo delle Reti Neurali Convoluzionali (CNN), la tecnologia utilizzata per l'analisi e il riconoscimento di immagini. Premio Turing 2018 (considerato il Nobel dell'informatica) è anche un convinto sostenitore dell'IA open-source.

Fonte: https://scholar.google.com/citations?user=WLN3QrAAAAAJ&hl=en

McCarthy, John

McCarthy, John (1927-2011). Informatico statunitense, propose il termine 'Artificial Intelligence' nel 1955 nell'ambito del *Dartmouth Summer Research Project*, poi svoltosi nel 1956. È inoltre ideatore del linguaggio di programmazione Lisp, ampiamente utilizzato nella ricerca sull'IA per diversi decenni. È annoverato tra i principali promotori delle prime ricerche organizzate nel campo dell'intelligenza artificiale.

Fonte: McCarthy J., Minsky M. L., Rochester N., Shannon C. E. (2006), A proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955, AI magazine, 27(4), pp.12-12

Meta

Azienda tecnologica fondata da Mark Zuckerberg, precedentemente nota come Facebook. Nel campo dell'IA, è uno dei principali sviluppatori di modelli linguistici open-source (la famiglia Llama) e ha destinato ingenti risorse all'intelligenza artificiale per il metaverso, i social media e la realtà aumentata.

Fonte: https://about.meta.com/

NVIDIA

Azienda americana diventata una delle più importanti al mondo grazie alle sue GPU (processori grafici) che sono indispensabili per addestrare e far funzionare i modelli di intelligenza artificiale. Le GPU NVIDIA sono una delle principali infrastrutture hardware su cui vengono eseguiti molti modelli di IA di grandi dimensioni, si potrebbero definire una sorta di 'motore' su cui girano praticamente tutti i grandi modelli di IA attuali.

Fonte: https://www.nvidia.com/

OpenAI

Laboratorio di ricerca sull'intelligenza artificiale fondato nel 2015, creatore di ChatGPT, GPT-4, DALL-E e Sora. Nata come organizzazione no-profit con la missione di sviluppare un'IA orientata alla sicurezza e al beneficio pubblico per l'umanità, si è poi evoluta in un modello ibrido con una sussidiaria a scopo di lucro. È uno degli attori di maggiore visibilità nel panorama attuale dell'IA generativa.

Fonte: https://openai.com/

Turing, Alan

Matematico, crittografo e teorico della computazione britannico (1912-1954). Contribuì, nell'ambito delle attività di *Bletchley Park* durante la Seconda guerra mondiale, alla decrittazione dei codici tedeschi, tra cui Enigma, è annoverato tra i principali fondatori dell'informatica e dell'intelligenza artificiale. Il suo celebre 'Test di Turing' ha posto le basi per la riflessione sull'intelligenza delle macchine.

Fonte: Bolognesi F. (2024), Intelligenza artificiale, Istruzioni per l'uso, EPC Editore

Turing Test

Un test proposto nel 1950 dal matematico Alan Turing per valutare se una macchina possa essere considerata 'intelligente'. Formulato come *imitation game* prevede che se un valutatore umano, conversando con una macchina e un essere umano senza vederli, non riesce a distinguerli, allora la macchina ha superato il test. È uno dei concetti fondativi dell'IA, anche se oggi è considerato insufficiente come unico criterio di intelligenza.

Fonte: Turing A. M. (1987), Computing machinery and intelligence (1950), Mind, 59(236), pp.33-60

5. Infrastruttura tecnologica

5.1 Componenti e architetture dei modelli di intelligenza artificiale

Quest'area della tassonomia permette di comprendere i meccanismi che compongono la macchina, descrivendone le componenti e le architetture che permettono ai modelli di intelligenza artificiale di funzionare. Nei precedenti paragrafi è stato presentato cosa fa l'IA e chi la sviluppa, in questa parte del documento si esplora come tali obiettivi vengono raggiunti: quali strutture matematiche e computazionali rendono possibile l'apprendimento automatico, la generazione di testo, il riconoscimento di immagini e tutte le altre capacità che caratterizzano i sistemi intelligenti moderni.

Il punto di partenza sono le reti neurali artificiali (ANN), modelli computazionali ispirati al funzionamento del cervello umano, composte da neuroni artificiali organizzati in strati che collaborano per elaborare informazioni. In questo paragrafo sarà possibile comprendere come queste reti 'imparino' aggiustando i pesi delle connessioni durante l'addestramento, e come la loro evoluzione abbia prodotto architetture sempre più sofisticate: dalle reti neurali convoluzionali (CNN), specializzate nell'elaborazione di immagini e alla base della computer vision, alle reti neurali ricorrenti (RNN), progettate per gestire dati sequenziali come il linguaggio e le serie temporali.

Si passa successivamente a descrivere l'architettura Transformer: la rivoluzione tecnica del 2017 che ha reso possibili i grandi modelli linguistici odierni. Il meccanismo di attenzione permette di fatto a questi modelli di analizzare tutte le parole di un testo simultaneamente, cogliendo relazioni anche tra elementi molto distanti. Accanto ai Transformer, vengono approfonditi concetti come gli *embedding*, che traducono parole e concetti in rappresentazioni numeriche, le reti generative avversarie (GAN), e il ruolo dei parametri come misura della complessità e della 'conoscenza' di un modello.

Sebbene i contenuti di questo paragrafo abbiano un carattere più tecnico, le definizioni sono state elaborate per essere accessibili anche a lettori senza una formazione specialistica, offrendo analogie e spiegazioni che rendono comprensibili anche i concetti più complessi. Al termine della lettura, si sarà acquisita una comprensione solida delle fondamenta architetturali su cui si regge l'intera industria dell'intelligenza artificiale.

ANN (Artificial Neural Network)

Il meccanismo di attenzione è una tecnica utilizzata nelle reti neurali che consente al modello di concentrarsi sulle parti più importanti dell'informazione in ingresso, ignorando quelle meno rilevanti. Può essere paragonato all'attenzione umana: quando leggiamo una frase, non attribuiamo lo stesso peso a ogni singola parola, ma ci focalizziamo su quelle chiave per capire il significato.

Fonte: Bahdanau D. (2014), Neural Machine Translation by Jointly Learning to Align and Translate, arXiv preprint arXiv:1409.0473

Artificial Neuron

Il Neurone artificiale è l'unità fondamentale delle reti neurali artificiali, ispirata ai neuroni del cervello umano. Riceve segnali in ingresso (input), li elabora attraverso un calcolo matematico e produce un risultato in uscita (output). Le unità sono connesse tra loro tramite pesi numerici che determinano l'intensità delle connessioni; l'insieme di tali connessioni costituisce una rete i cui parametri vengono ottimizzati durante l'addestramento.
La sua funzione è trasformare e filtrare le informazioni in modo che, collegando molti neuroni tra loro in una rete, il sistema possa riconoscere schemi nei dati e apprendere a svolgere compiti complessi, come classificare immagini, prevedere valori o generare testo.

Fonte: Alzahrani A., Parker C. (2020), Neuromorphic Circuits with Neural Modulation Enhancing the Information Content of Neural Signaling, Proceedings of International Conference on Neuromorphic Systems

Attention Mechanism

Il meccanismo di attenzione è una tecnica utilizzata nelle reti neurali che consente al modello di concentrarsi sulle parti più importanti dell'informazione in ingresso, ignorando quelle meno rilevanti. Può essere paragonato all'attenzione umana: quando leggiamo una frase, non attribuiamo lo stesso peso a ogni singola parola, ma ci focalizziamo su quelle chiave per capire il significato.

Fonte: Bahdanau D. (2014), Neural Machine Translation by Jointly Learning to Align and Translate, arXiv preprint arXiv:1409.0473

CNN (Reti Neurali Convoluzionali)

Tipo specializzato di rete neurale particolarmente adatto all'elaborazione di dati visivi come immagini e video. Si può paragonare al modo con cui funziona il sistema visivo umano: analizza l'immagine partendo dai dettagli più piccoli (bordi, colori, forme) per poi riconoscere strutture sempre più complesse (occhi, volti, oggetti interi). Tali reti rappresentano la tecnologia alla base del riconoscimento facciale, dei filtri fotografici e della guida autonoma.

Fonte: LeCun Y. *et al.* (1998), Gradient-Based Learning Applied to Document Recognition, Proceedings of the IEEE

Embedding

Si tratta di una tecnica finalizzata a rappresentare concetti complessi (come parole, frasi o immagini) sotto forma di sequenze di numeri (vettori) all'interno di uno spazio multidimensionale. In questa rappresentazione matematica, elementi semanticamente simili vengono collocati in posizioni vicine tra loro: ad esempio, i vettori associati a 're' e 'regina' risulteranno più prossimi rispetto a 're' e 'automobile'.
Questa organizzazione numerica consente ai modelli di individuare relazioni e somiglianze tra concetti senza comprenderli nel senso umano del termine. Gli *embedding* costituiscono una componente fondamentale dei grandi modelli linguistici e dei sistemi di ricerca semantica.

Fonte: Bengio Y., Ducharme R., Vincent P. (2003), A Neural Probabilistic Language Model in Journal of Machine Learning Research, 3, pp.1137–1155

GAN (Generative Adversarial Network)

Architettura di intelligenza artificiale composta da due reti neurali che lavorano in opposizione tra loro al fine di generare contenuti realistici.
La prima rete è chiamata generatore e crea nuovi contenuti (ad esempio immagini) partendo da dati casuali.
La seconda rete viene detta discriminatore e ha il compito di valutare se il contenuto prodotto sia autentico oppure artificiale.
Il generatore cerca continuamente di 'ingannare' il discriminatore producendo risultati sempre più credibili, mentre il discriminatore migliora nel riconoscere quelli falsi. Questo processo di competizione reciproca porta gradualmente alla produzione di contenuti molto realistici.
Ad esempio, se il sistema viene addestrato su migliaia di fotografie di volti umani, il generatore impara a produrre nuovi volti sintetici sempre più realistici, mentre il discriminatore tenta di riconoscere quali immagini siano autentiche e quali generate.
Le GAN sono state tra le prime tecniche efficaci per la generazione di immagini sintetiche e sono alla base di molte applicazioni di *image generation* e *deepfake*.

Fonte: Goodfellow I. *et al.* (2014), Generative Adversarial Networks', NeurIPS

Parameter

I 'parametri' di un modello di IA rappresentano l'informazione incorporata nel modello attraverso l'addestramento. Si tratta di milioni (o miliardi) di valori numerici interni che il modello apprende durante l'addestramento. Quando si afferma che un modello possiede 'centinaia di miliardi di parametri', significa che contiene altrettanti valori numerici che determinano il modo in cui elabora le informazioni e genera le risposte.
I parametri non costituiscono una memoria nel senso tradizionale; piuttosto, incorporano configurazioni numeriche che riflettono le regolarità statistiche presenti nei dati di addestramento.

Fonte: Goodfellow I., Bengio Y., Courville A., Bengio Y. (2016), Deep learning (vol.1, n.2, pp.1-800), Cambridge, MIT Press

RNN (Recurrent Neural Network)

La Rete neurale ricorrente è un tipo di rete neurale progettata per elaborare sequenze di dati (come frasi, serie temporali o brani musicali) mantenendo una sorta di memoria delle informazioni precedenti. Questo la rende adatta a compiti in cui l'output in un determinato momento dipende anche dagli elementi precedenti della sequenza, come nella traduzione automatica, nel riconoscimento vocale e nell'analisi di serie temporali.

Fonte: Rumelhart D. E. *et al.* (1986), Learning Internal Representations by Error Propagation, Nature

Transformer

Architettura di rete neurale introdotta nel 2017 con il paper *Attention Is All You Need*, che ha segnato un cambiamento rilevante nello sviluppo dei modelli per l'elaborazione di sequenze, in particolare nel linguaggio naturale.
Il suo elemento distintivo è l'uso sistematico del meccanismo di attenzione, architettura di rete neurale introdotta nel 2017 che permette ai modelli di analizzare un'intera frase o sequenza tutta insieme, anziché parola per parola al fine di identificare quali parti del testo sono più importanti tra loro. È la tecnologia alla base dei moderni modelli linguistici come GPT.

Fonte: Vaswani A. *et al.* (2017), Attention Is All You Need, NeurIPS

5.2 Metodi di apprendimento e tecniche di addestramento

A differenza del paragrafo precedente che ha illustrato le strutture interne dei modelli di IA, questa sezione si concentra su come questi modelli imparano. In questa parte del lavoro si evidenziano metodi e tecniche di addestramento che trasformano un modello 'vuoto' in un sistema capace di riconoscere pattern, fare previsioni e generare contenuti. Comprendere questi processi è fondamentale per capire sia le potenzialità sia i limiti dell'intelligenza artificiale.

In quest'area della tassonomia rientrano i principali paradigmi del *Machine Learning*: l'apprendimento supervisionato, in cui il modello impara da esempi etichettati; l'apprendimento non supervisionato, in cui individua autonomamente strutture latenti nei dati; e l'apprendimento per rinforzo, basato su un meccanismo di prova, errore e ricompensa. A questi si affianca l'RLHF (*Reinforcement Learning from Human Feedback*), tecnica utilizzata per allineare il comportamento dei modelli linguistici alle preferenze e alle aspettative umane.

Inoltre, il paragrafo approfondisce il ciclo di vita dell'addestramento, distinguendo tra pre-training – fase iniziale di acquisizione di conoscenza generale su larga scala – e fine-tuning, processo di specializzazione su compiti specifici. Vengono analizzate anche le principali tecniche di ottimizzazione, come la *backpropagation* e il *transfer learning*, nonché le criticità tipiche dell'addestramento, tra cui l'*overfitting*.

Per finire vengono descritti i compiti fondamentali del *machine learning*, quali classificazione, *clustering*, *data mining, pattern recognition*, analisi predittiva e alberi decisionali.

5.2.1 Paradigmi fondamentali del machine learning

Deep Learning

L'apprendimento profondo è una branca avanzata del *Machine Learning* basata su reti neurali artificiali con molti strati sovrapposti (da qui il termine 'profondo'). Questa profondità permette al sistema di apprendere rappresentazioni dei dati sempre più astratte e sofisticate.
Queste reti vengono addestrate su grandi quantità di esempi etichettati o non etichettati. Ogni volta che effettuano una previsione, confrontano il risultato ottenuto con quello corretto e si autocorreggono gradualmente, modificando i collegamenti interni per ridurre progressivamente l'errore. Ripetendo questo processo migliaia o milioni di volte, migliorano progressivamente le prestazioni.
La presenza di molteplici strati uno sovrapposto all'altro permette alla rete di analizzare le informazioni a livelli diversi: prima riconosce elementi semplici (per esempio linee o suoni), poi combina questi elementi per individuare forme, parole o concetti più complessi.
Lo sviluppo del *Deep Learning* è stato reso possibile dalla disponibilità di grandi quantità di dati e dall'aumento della capacità computazionale (in particolare tramite GPU). È la tecnologia alla base di numerosi avanzamenti dell'IA, dal riconoscimento vocale alla traduzione automatica, dalla generazione di immagini ai modelli linguistici.

Fonte: LeCun Y. *et al.* (2015), Deep Learning, Nature

RL (Reinforcement Learning)

L'Apprendimento per rinforzo è un approccio al *Machine Learning* ispirato al modo in cui apprendono gli animali e gli esseri umani: attraverso un meccanismo di prova ed errore. Un agente IA compie azioni in un ambiente e riceve ricompense (quando fa bene) o penalità (quando sbaglia), imparando progressivamente la strategia migliore. È la tecnica usata per l'addestramento di sistemi IA che giocano a videogiochi, guidano veicoli autonomi e ottimizzano processi industriali. Le ricompense sono semplicemente punteggi numerici che indicano all'agente se una scelta è stata adeguata o meno, aiutandolo a migliorare nel tempo.

Fonte: Sutton R. S., Barto, A. G. (2018), Reinforcement Learning: An Introduction, 2nd edition, MIT Press

RLHF (Reinforcement Learning from Human Feedback)

L'apprendimento per rinforzo basato sul feedback umano è una tecnica chiave per rendere i modelli di IA più utili e sicuri: valutatori umani giudicano le risposte del modello, e queste valutazioni vengono usate per addestrare il modello stesso a produrre risposte migliori, più precise e più allineate alle aspettative degli utenti. È la tecnica che ha reso molti LLM efficaci nel dialogo. È opportuno considerare che questo processo avviene durante la fase di addestramento del modello, prima della sua messa in uso, e non durante le singole conversazioni con gli utenti.

Fonte: Ouyang L., Wu J., Jiang X., Almeida, D., Wainwright C., Mishkin P., Lowe, R. (2022), Training Language Models to Follow Instructions with Human Feedback, Advances in Neural Information Processing Systems, 35, pp.27730-27744

Supervised Learning

L'apprendimento supervisionato costituisce il metodo di addestramento più comune: il modello viene addestrato su un set di dati in cui ogni esempio è accompagnato dalla risposta corretta (etichetta). Il sistema apprende osservando molti casi in cui conosce già la soluzione, e prova progressivamente a riprodurla da solo.
Per analogia, è simile a un processo in cui le soluzioni corrette sono note in anticipo e vengono utilizzate per correggere gli errori di previsione.
Ad esempio, se il modello viene addestrato su migliaia di e-mail già classificate come 'spam' o 'non spam', i suoi parametri vengono ottimizzati in modo da associare determinate caratteristiche del testo alla relativa etichetta; tali associazioni possono poi essere applicate alla classificazione di nuovi messaggi non presenti nel set di addestramento.

Fonte: Russell S., Norvig P. (2020), Artificial Intelligence: A Modern Approach, 4th edition

Unsupervised Learning

L'apprendimento non supervisionato è una tipologia di algoritmo progettata per sviluppare un paradigma di apprendimento autonomo, non guidato da etichette o risposte predefinite. La sua logica si fonda sull'idea che, pur in presenza di dati caratterizzati da una ricca struttura intrinseca, le informazioni di riferimento corrette (etichette target) siano generalmente assenti, pur essendo comunque definiti criteri matematici o funzioni obiettivo per l'ottimizzazione del modello.
Ne consegue che la maggior parte delle informazioni apprese dall'algoritmo deve emergere direttamente dalla struttura interna dei dati in input, piuttosto che dall'applicazione di una conoscenza specificamente orientata a un compito pratico.

Fonte: Tyagi K., Rane C., Sriram R., Manry M. (2022), Unsupervised Learning, In Artificial Intelligence and Machine Learning for Edge Computing, pp.33-52, Academic Press

5.2.2 Ciclo di vita e tecniche di addestramento

Pre-training

La fase iniziale di addestramento in cui un modello di IA viene istruito su grandi quantità di dati generici (ad esempio miliardi di pagine web, libri e articoli) per modellare regolarità statistiche e strutture linguistiche presenti nei dati. Dopo questa fase, il modello può essere ulteriormente affinato (*fine-tuning*) per compiti specifici.

Fonte: Devlin J. *et al.* (2018), BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, arXiv preprint
Brown T. B. *et al.* (2020), Language Models are Few-Shot Learners, NeurIPS

Fine-tuning

Il *fine-tuning* è il processo di adattamento e perfezionamento di un modello di IA già addestrato su dati generici, utilizzando un insieme più piccolo e mirato di dati relativi a un compito specifico.
Il modello dispone già di parametri ottimizzati su grandi quantità di dati, che catturano regolarità statistiche di carattere generale. Con il *fine-tuning*, tali parametri vengono ulteriormente aggiornati su dati più specifici, in modo da migliorare le prestazioni del modello in un determinato ambito applicativo.
Per analogia funzionale, si può paragonare a una fase di specializzazione successiva: una base generale viene adattata e orientata verso un dominio più circoscritto.

Fonte: Devlin J., *et al.* (2018), BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding, arXiv preprint

Backpropagation

La *backpropagation* è il procedimento matematico che consente a una rete neurale di correggere progressivamente i propri parametri quando l'output prodotto è diverso da quello atteso. Dopo ogni previsione, il risultato ottenuto viene confrontato con quello corretto, e la differenza (errore) viene propagata all'indietro attraverso i livelli della rete per capire quali collegamenti interni hanno contribuito maggiormente a sbagliare.
In base a queste informazioni, il sistema modifica leggermente i pesi delle connessioni interne, in modo da ridurre l'errore nelle previsioni successive.
Ripetendo questo processo migliaia o milioni di volte, la rete migliora progressivamente le proprie prestazioni.

Fonte**:** Rumelhart D. E. *et al.* (1986), Learning Representations by Back-Propagating Errors, Nature

Overfitting

Il sovra-adattamento (*overfitting*) è una condizione che può verificarsi durante l'addestramento di un modello di IA quando questo si adatta eccessivamente ai dati di training, includendo anche il rumore, le fluttuazioni casuali o caratteristiche non rappresentative della distribuzione generale. In tali casi, il modello mostra prestazioni elevate sui dati di addestramento ma una ridotta capacità di generalizzazione su dati nuovi.

Fonte**:** Goodfellow I. *et al*. (2016), Deep Learning, MIT Press

Transfer Learning

L'apprendimento per trasferimento è una tecnica in cui un modello già addestrato su un compito viene riutilizzato come punto di partenza per affrontarne un altro, simile e correlato. In questo modo, parte delle rappresentazioni e dei parametri già ottimizzati vengono mantenuti e ulteriormente aggiornati. Questo approccio consente di risparmiare enormi quantità di tempo, dati e risorse computazionali.
Per analogia funzionale, il processo può essere paragonato al passaggio di un musicista dalla formazione classica al jazz: non si riparte da zero, ma si riutilizzano strutture e abilità già consolidate, adattandole a un nuovo contesto.

Fonte: Pan S. J., Yang Q. (2010), A Survey on Transfer Learning, IEEE Transactions on Knowledge and Data Engineering

5.2.3 Compiti tipici del machine learning

Classificazione

Un compito fondamentale del *Machine Learning* in cui il modello viene addestrato ad assegnare un'etichetta o una categoria a un dato in ingresso sulla base di esempi etichettati. Ad esempio: classificare un'e-mail come spam o non spam, riconoscere se un'immagine contiene un gatto o un cane, o determinare se una transazione bancaria è fraudolenta. Una volta completato l'addestramento, il modello applica le regole apprese dai dati di training per classificare nuovi esempi.

Fonte: Gordon A. D. (1999), Classification, CRC Press

Clustering

Il clustering è una tecnica di apprendimento non supervisionato che organizza automaticamente i dati in gruppi (cluster) sulla base di criteri di similarità definiti matematicamente, senza l'utilizzo di etichette predefinite. Per analogia funzionale, il processo può essere paragonato all'organizzazione di una libreria disordinata in cui i libri vengono raggruppati per affinità di contenuto, senza che le categorie siano stabilite in anticipo. È impiegato in diversi ambiti applicativi, tra cui la segmentazione della clientela, l'organizzazione di documenti e il rilevamento di anomalie.

Fonte: Britannica: https://www.britannica.com/topic/cluster-analysis

Data Mining

Data Mining significa cercare informazioni nascoste nei dati. Si tratta di un processo di esplorazione e analisi di grandi quantità di dati finalizzato all'individuazione di pattern, relazioni e informazioni significative e non immediatamente evidenti. Con tale processo vengono combinate diverse tecniche statistiche, algoritmi di *machine learning* e strumenti di gestione dei dati per produrre informazioni e modelli utili a supportare processi decisionali.
Il *data mining* è utilizzato, ad esempio, per identificare tendenze nei comportamenti dei clienti, rilevare frodi, prevedere, in termini probabilistici, fenomeni futuri o individuare correlazioni tra variabili. Per analogia funzionale, il processo può essere paragonato al setacciare grandi quantità di materiale per individuare elementi rilevanti, secondo criteri definiti.

Fonte: Han J. *et al.* (2011), Data Mining: Concepts and Techniques. 3rd edition, Morgan Kaufmann

Decision Tree

L'Albero decisionale è un metodo di *Machine Learning* che struttura il processo di classificazione o regressione secondo una forma gerarchica ad albero: partendo dalla radice il modello verifica una serie di condizioni sui valori delle variabili. Un esempio è: "Il reddito è superiore a 30.000€?", "L'età è maggiore di 25?". In base alle risposte, il percorso segue un ramo diverso fino a raggiungere un nodo finale, che rappresenta la decisione o la previsione.
La struttura comprende nodi interni, che definiscono le condizioni di suddivisione dei dati, e nodi terminali (foglie), che forniscono la classificazione o la stima finale.

Fonte: Charbuty B., Abdulazeez A. (2021), Classification Based on Decision Tree Algorithm for Machine Learning. Journal of Applied Science and Technology Trends, 2(01), pp.20-28

Pattern Recognition

Il riconoscimento di schemi *(Pattern Recognition*) è un ambito interdisciplinare che studia metodi per identificare strutture ricorrenti e regolarità nei dati, spesso mediante tecniche statistiche e di *Machine Learning*. Quando il telefono riconosce il vostro viso, un'app identifica una canzone o un sistema rileva una frode bancaria, sta applicando la *pattern recognition*.
Nel riconoscimento facciale, l'immagine del volto viene rappresentata come un insieme di valori numerici derivati dalle configurazioni dei pixel. Queste rappresentazioni numeriche vengono confrontate con quelle già archiviate nel sistema, calcolando quanto sono simili tra loro; se la somiglianza supera una certa soglia, il sistema le considera corrispondenti.

Fonte: Howard W.R. (2007), Pattern Recognition and Machine Learning, Kybernetes, vol. 36 n.2, pp.275-275

Predictive Analytics

L'analisi predittiva consente di generare previsioni a partire dai dati. Essa consiste nell'uso di dati storici, tecniche statistiche e modelli di intelligenza artificiale per stimare in termini probabilistici, possibili eventi futuri. Viene impiegata in ambiti molto diversi: dalla previsione del meteo alla stima della domanda di prodotti, dalla valutazione del rischio finanziario alla medicina preventiva.
Attraverso tali modelli vengono identificate relazioni e regolarità nei dati storici, al fine di generare stime probabilistiche relative a eventi futuri di interesse.

Fonte: https://www.ibm.com/it-it/think/topics/predictive-analytics
Shmueli G., Koppius O. R. (2011), Predictive analytics in information systems research1, MIS Quarterly, 35(3), pp.553-572

5.3 Linguaggi, strumenti e framework per lo sviluppo

Questo paragrafo è funzionale a illustrare gli strumenti pratici che rendono possibile lo sviluppo, l'implementazione e l'utilizzo dei sistemi di intelligenza artificiale. Mentre le categorie precedenti hanno esplorato i concetti, le architetture e i metodi dell'AI, qui si descrive l'infrastruttura software: i linguaggi di programmazione, le piattaforme di sviluppo, i protocolli di comunicazione e gli ambienti di lavoro che traducono la teoria in applicazioni concrete.

Al centro della sezione della tassonomia si trova Python, il linguaggio di programmazione che domina il campo dell'intelligenza artificiale e della scienza dei dati. Grazie alla sua sintassi accessibile e alla ricchezza delle sue librerie specializzate, Python è diventato lo strumento d'elezione per ricercatori e sviluppatori di tutto il mondo. Accanto a Python, si trovano le API (*Application Programming Interface*), i meccanismi che permettono a software diversi di comunicare tra loro e che rappresentano il collante dell'ecosistema

tecnologico: è grazie alle API, ad esempio, che un'applicazione può integrare le capacità di un modello linguistico come ChatGPT, Claude o Gemini.

La sezione presenta inoltre gli strumenti di sviluppo assistito dall'IA che stanno trasformando il lavoro dei programmatori: GitHub Copilot e Cursor, editor intelligenti che suggeriscono e generano codice in tempo reale, rappresentando una nuova generazione di ambienti di sviluppo in cui l'IA diventa un collaboratore attivo.

Completano il paragrafo le piattaforme No Code AI, che democratizzano l'accesso all'intelligenza artificiale permettendo anche a persone senza competenze di programmazione di creare soluzioni basate sull'IA, e Ollama, che consente di eseguire modelli generativi direttamente sul proprio computer in modo privato e autonomo. Questa categoria è particolarmente utile per chi desidera comprendere il panorama degli strumenti disponibili e orientarsi tra le diverse opzioni per iniziare a lavorare concretamente con l'intelligenza artificiale.

API (Application Programming Interface)

L'interfaccia di programmazione delle applicazioni è un insieme di regole che permette a programmi diversi di comunicare tra loro. Un'API definisce le modalità con cui un software può richiedere funzioni o dati a un altro software, sia in locale sia attraverso una rete.
Nel campo dell'intelligenza artificiale, le API rappresentano il meccanismo attraverso cui sviluppatori e ricercatori possono integrare modelli già addestrati all'interno dei propri sistemi, senza doverli installare o addestrare direttamente.
Permettono, ad esempio, a un'app meteo di ricevere informazioni aggiornate da un servizio centrale o a un sito web di ottenere una risposta generata da ChatGPT senza dover ospitare il modello al proprio interno.
Ad esempio, uno sviluppatore può scrivere un programma in Python che invia una richiesta a un modello di intelligenza artificiale tramite API. Il codice Python invia il testo da analizzare ai server del modello e riceve in risposta il risultato (come un *embedding*, una classificazione o un testo generato).
Per analogia funzionale, può essere descritta come un meccanismo di intermediazione che riceve una richiesta, la instrada al servizio competente e restituisce il risultato all'applicazione chiamante.

Fonte: Cambridge Dictionary: API; cfr. Treccani: https://www.treccani.it/enciclopedia/api_(Lessico-del-XXI-Secolo)/

Cursor

Editor di codice integrato con modelli di intelligenza artificiale che supporta lo sviluppo software attraverso suggerimenti contestuali, generazione automatica di funzioni e assistenza nella modifica del codice esistente. Rappresenta una nuova generazione di strumenti di sviluppo in cui modelli di IA sono integrati nel flusso di sviluppo per fornire suggerimenti e generare risposte a interrogazioni sul codice.
A differenza dei tradizionali editor con semplice autocompletamento, Cursor è progettato con un'integrazione nativa di modelli di AI: può analizzare il contesto dell'intero progetto, rispondere a domande sul codice e proporre modifiche strutturate, integrando funzionalità di assistenza automatizzata nel processo di programmazione.
Per analogia funzionale, può essere descritto come uno strumento di supporto che fornisce suggerimenti contestuali durante la scrittura del codice.

Fonte: https://www.cursor.so/
Chen M., Tworek J., Jun H., Yuan Q., Pinto H. P. D. O., Kaplan J., Zaremba W. (2021), Evaluating Large Language Models Trained on Code, arXiv preprint arXiv:2107.03374

GitHub Copilot

Strumento di intelligenza artificiale sviluppato da GitHub (Microsoft) e OpenAI che assiste i programmatori suggerendo codice in tempo reale mentre scrivono. Analizza il contesto del progetto proponendo automaticamente frammenti di codice o funzioni complete, con l'obiettivo di ridurre il tempo necessario alla scrittura del codice.

Fonte: https://github.com/features/copilot
Barke S., James M. B., Polikarpova N. (2022), Grounded Copilot: How Programmers Interact with Code-generating Models, (2022), CoRR arXiv, 2206

No Code AI Platforms

Piattaforme che permettono di creare e utilizzare modelli di intelligenza artificiale senza scrivere codice. Democratizzano l'accesso all'IA, rendendo possibile l'utilizzo di modelli di IA anche da parte di utenti con limitate competenze di programmazione, pur richiedendo competenze di dominio e comprensione dei dati. Ad esempio, un utente può caricare un dataset, scegliere l'obiettivo (classificazione o previsione) e la piattaforma automatizza le fasi di selezione, addestramento e validazione del modello. Esempi includono Google AutoML, Microsoft Azure Machine Learning, DataRobot e sono usati soprattutto in ambito aziendale.

Fonte: Kang M. *et al.* (2024), No-Code AI: Concepts and Applications in Machine Learning, Visualization and Cloud Platforms

Ollama

Ollama è una piattaforma open-source che consente di eseguire localmente modelli di intelligenza artificiale generativa, senza ricorrere a servizi cloud durante la fase di inferenza. Supporta l'esecuzione di modelli open-source sul proprio dispositivo, compatibilmente con i requisiti hardware richiesti. L'esecuzione locale può offrire un maggiore controllo infrastrutturale e gestionale dei dati, aspetto rilevante in contesti che richiedono riservatezza o autonomia operativa.

Fonte: https://ollama.com/

Python

Uno dei linguaggi di programmazione più utilizzati nel campo dell'Intelligenza Artificiale e della scienza dei dati. Nel campo dell'IA, Python funge da ecosistema di sviluppo, grazie all'ampia disponibilità di librerie e strumenti integrati. Consente di elaborare dati, addestrare modelli, generare *embedding*, costruire pipeline di analisi e integrare servizi esterni tramite API.
La diffusione di Python nell'IA non deriva solo dalla semplicità sintattica ma dalla concentrazione di librerie scientifiche e framework che lo hanno reso uno standard di riferimento nella sperimentazione e nello sviluppo di modelli di *machine learning* e *deep learning*.
Principali librerie:

- analisi dati e calcolo numerico: NumPy, Pandas, SciPy;
- visualizzazione: Matplotlib, Seaborn, Plotly;
- machine learning: Scikit-learn, XGBoost, LightGBM;
- deep learning: TensorFlow, PyTorch, Keras;
- NLP: NLTK, SpaCy, Transformers, Sentence-transformers;
- integrazione API e servizi: Requests, OpenAI, FastAPI.

Fonte: VanderPlas J. (2016), Python Data Science Handbook: Essential Tools for Working with Data, O'Reilly Media, Inc.

5.4 Hardware per l'intelligenza artificiale

L'intelligenza artificiale non è solo software: dietro ogni modello linguistico, ogni sistema di riconoscimento e ogni generatore di immagini c'è un'infrastruttura hardware essenziale per il funzionamento. Questa sezione illustra le componenti che alimentano la rivoluzione dell'IA, dai processori specializzati nei grandi data center ai chip miniaturizzati integrati nei nostri smartphone.

Il protagonista indiscusso è la GPU (*Graphics Processing Unit*), il processore nato per i videogiochi che si è rivelato perfetto per l'addestramento delle reti neurali grazie alla sua capacità di eseguire migliaia di calcoli in parallelo. Si comprenderà perché NVIDIA, il principale produttore di GPU per l'IA, sia diventata una delle aziende più importanti e valorizzate al mondo. Accanto alle GPU, la sezione presenta le TPU (*Tensor Processing Unit*), i chip sviluppati da Google specificamente per ottimizzare i carichi di lavoro dell'intelligenza artificiale, e le NPU (*Neural Processing Unit*), processori a basso consumo energetico progettati per portare l'IA direttamente nei dispositivi personali.

Un concetto chiave della sezione della tassonomia è l'Edge AI, ovvero l'esecuzione di algoritmi di intelligenza artificiale direttamente sui dispositivi dell'utente anziché su server remoti nel cloud. Questo approccio offre vantaggi significativi in termini di velocità di risposta e tutela della privacy, ed è alla base di funzionalità quotidiane come il riconoscimento facciale sullo smartphone. La sezione esplora inoltre la categoria più generale degli acceleratori hardware per l'IA, chip specializzati che offrono prestazioni superiori rispetto ai processori generici.

Chiude l'area degli hardware la tecnologia computazionale basata sulla meccanica quantistica che promette di affrontare problemi oggi irrisolvibili e che potrebbe, in futuro, accelerare in modo esponenziale le capacità dell'intelligenza artificiale.

Questa categoria della trattazione offre una prospettiva fondamentale: comprendere che i progressi dell'IA dipendono tanto dall'innovazione software quanto dall'evoluzione dell'hardware, e che i limiti fisici dei processori rappresentano una delle sfide cruciali per il futuro del settore.

AI HW Accelerator

Un acceleratore hardware per l'IA è un dispositivo progettato per eseguire in modo efficiente le operazioni computazionali tipiche dei modelli di *machine learning* e *deep learning*, come il calcolo su matrici e vettori. Un AI HW Accelerator è un chip speciale fatto per far funzionare l'IA più velocemente. Rispetto a un processore generico è uno strumento fortemente specializzato, un acceleratore IA estremamente efficiente nel suo compito specifico.

Fonte: IBM Think: https://www.ibm.com/think/topics/ai-accelerator

Edge AI

Edge AI significa far funzionare l'IA direttamente sul proprio dispositivo (telefono, telecamera, sensore) piuttosto che procedere su internet. L'Edge AI consiste nell'eseguire modelli e algoritmi di intelligenza artificiale su dispositivi situati ai margini della rete, evitando il ricorso a infrastrutture centralizzate come i data center cloud.
Il vantaggio è duplice: risposte più rapide (senza dover inviare dati a un server lontano) e maggiore privacy (i dati restano sul dispositivo). Un esempio è il riconoscimento facciale sullo smartphone, che funziona anche senza connessione Internet.

Fonte: López M. *et al.* (2020), Edge AI: A Survey, IEEE Internet of Things Journal
Singh R., Gill S. S. (2023), Edge AI: a Survey, Internet of Things and Cyber-Physical Systems, 3, pp.71-92

GPU (Graphics Processing Unit)

L'unità di elaborazione grafica è un processore sviluppato originariamente per l'elaborazione grafica, ma divenuto uno dei componenti hardware centrali nello sviluppo dell'intelligenza artificiale moderna. A differenza delle CPU (*Central Processing Unit*) tradizionali — progettate per gestire operazioni generali in modo prevalentemente sequenziale — le GPU sono ottimizzate per l'elaborazione parallela e possono eseguire simultaneamente un elevato numero di operazioni. Questa caratteristica le rende particolarmente adatte all'addestramento di modelli di *machine learning* e *deep learning*, che comportano un'intensa attività di calcolo, in particolare su matrici e vettori. Tra i principali produttori di GPU per applicazioni di IA figurano aziende come NVIDIA e AMD.

Fonte: NVIDIA Glossary: GPU
Owens J. D., Houston M., Luebke D., Green S., Stone J. E., Phillips, J. C. (2008), GPU Computing, Proceedings of the IEEE, 96(5), pp.879-899

NPU (Neural Processing Unit)

L'Unità di elaborazione neurale è un tipo di processore progettato per eseguire in modo efficiente le operazioni computazionali tipiche dei modelli basati su reti neurali.
A differenza delle GPU, spesso impiegate per carichi di lavoro ad alte prestazioni come l'addestramento di modelli di grandi dimensioni, le NPU sono generalmente progettate per l'integrazione in dispositivi come computer, smartphone e sistemi embedded.
Ciò consente l'esecuzione locale di modelli di IA sul dispositivo stesso, riducendo la necessità di ricorrere a servizi cloud per l'inferenza. Si tratta di chip dedicati ottimizzati per l'elaborazione a bassa latenza (ossia con tempi di risposta ridotti) e per l'efficienza energetica in contesti *edge*.

Fonte: Fiack L., Rodriguez L., Miramond B. (2015), Hardware Design of a Neural Processing Unit for Bio-inspired Computing, In 2015 IEEE 13th International New Circuits and Systems Conference (NEWCAS) (pp. 1-4), IEEE

Quantum Computing

Il calcolo quantistico è una tecnologia computazionale basata sui principi della meccanica quantistica che promette di risolvere problemi troppo complessi per i computer tradizionali. In ambito di ricerca, si ipotizza che possa contribuire ad accelerare alcune procedure di ottimizzazione o simulazione rilevanti anche per il *machine learning*, nonché affrontare problemi che, con le tecnologie attuali, richiederebbero risorse computazionali proibitive (ad esempio nella simulazione di sistemi molecolari complessi).
Sebbene sia oggetto di intensa ricerca e sviluppo, si tratta di una tecnologia ancora emergente, con applicazioni pratiche limitate e non ancora diffuse nel campo dell'intelligenza artificiale.

Fonte: Biamonte J. *et al.* (2017), Quantum Machine Learning, Nature

TPU (Tensor Processing Unit)

Le unità di elaborazione sensoriale sono chip sviluppati da Google per eseguire in modo efficiente le operazioni computazionali tipiche delle reti neurali. A differenza delle GPU — nate per l'elaborazione grafica e successivamente adattate all'IA — le TPU sono state progettate fin dall'inizio per supportare in modo dedicato l'addestramento e l'inferenza di modelli di intelligenza artificiale in contesti data center.
Le TPU sono utilizzate principalmente nell'infrastruttura cloud di Google e rappresentano un esempio di hardware specializzato per l'IA.

Fonte: Jouppi N. P., Young C., Patil N., Patterson D., Agrawal G., Bajwa R., Yoon D. H. (2017), In-Datacenter Performance Analysis of a Tensor Processing Unit

5.5 Valutazione e benchmarking dei modelli

Questa sezione affronta un aspetto cruciale e spesso sottovalutato dell'intelligenza artificiale: la valutazione delle prestazioni, la trasparenza dei processi decisionali e la verifica dell'affidabilità dei risultati.

Il punto di partenza è il benchmarking, ovvero l'insieme delle tecniche e dei test standardizzati utilizzati per misurare e confrontare le capacità dei modelli di IA. In questa parte della trattazione si comprenderà come funzionano questi test a cui i modelli vengono sottoposti e perché sono essenziali per orientarsi tra le affermazioni, spesso enfatiche, che accompagnano il lancio di ogni nuovo modello sul mercato. I benchmark coprono ambiti diversi, dalla comprensione del testo al ragionamento matematico, dalla generazione di codice alla capacità di seguire istruzioni complesse.

La sezione affronta poi una delle sfide più rilevanti dell'IA contemporanea: il problema della 'scatola nera' (Black-Box). Molti modelli avanzati producono risultati eccellenti, ma non sono in grado di spiegare come ci arrivano: una situazione problematica in ambiti critici come la medicina, la giustizia o la finanza, dove la trasparenza delle decisioni non è un optional ma una necessità. In risposta a questa sfida, sono spiegati concetti come *Explainable* AI (XAI), il campo di ricerca dedicato a rendere comprensibili e interpretabili le decisioni dei sistemi di intelligenza artificiale.

Completano il paragrafo le tecniche di verifica automatizzata delle informazioni generate dai modelli, un aspetto di crescente importanza nell'era della disinformazione e dei contenuti generati artificialmente. Questa parte del lavoro fornisce gli strumenti concettuali per valutare criticamente le prestazioni e l'affidabilità dei sistemi di IA, andando oltre le promesse di marketing e adottando uno sguardo informato e consapevole.

Benchmarking

L'insieme delle tecniche e dei test standardizzati utilizzati per misurare e confrontare le prestazioni dei modelli di intelligenza artificiale. Per analogia funzionale, i benchmark possono essere paragonati a esami standardizzati: i modelli vengono sottoposti a compiti specifici — ad esempio elaborazione del linguaggio naturale, risoluzione di problemi matematici o generazione di codice — e valutati secondo metriche e criteri predefiniti.

Fonte: Todorov I., Penchev V. (2024), AI-Benchmarks and Datasets for LLM Evaluation, arXiv preprint: https://arxiv.org/pdf/2412.01020v1

Black-Box

La Scatola nera si riferisce a un sistema interno dell'intelligenza artificiale di cui non è possibile osservare o comprendere il funzionamento interno. È possibile conoscere gli input (i dati in ingresso) e gli output prodotti, ma non il processo attraverso cui tali risultati vengono generati. L'opacità può derivare sia dalla complessità matematica del modello (come nel caso delle reti neurali profonde), sia dalla mancanza di accesso ai parametri interni o ai dati di addestramento. Tale condizione solleva questioni di trasparenza, responsabilità e verificabilità, soprattutto in ambiti ad alto impatto come la medicina o il sistema giudiziario.

Fonte: Manakul P., Liusie A., Gales M. J. (2023), Selfcheckgpt: Zero-resource Black-box Hallucination Detection for Generative Large Language Models, arXiv preprint arXiv:2303.08896

Fact Checking

Il *fact checking* è una pratica sistematica di verifica delle affermazioni finalizzata ad accertarne l'accuratezza attraverso il confronto con evidenze documentate, dati verificabili e fonti attendibili. Non si limita alla semplice correzione di errori, ma rappresenta un processo epistemico volto a valutare la qualità dell'informazione e la solidità delle affermazioni nel dibattito pubblico.
In ambito digitale e nell'ecosistema dell'intelligenza artificiale, il *fact checking* assume una rilevanza crescente: può essere supportato o parzialmente automatizzato mediante strumenti computazionali che confrontano dichiarazioni testuali con basi informative strutturate, contribuendo a contrastare la disinformazione e a verificare l'accuratezza dei contenuti generati automaticamente.

Fonte: Guo Z., Schlichtkrull M., Vlachos A. (2022), A Survey on Automated Fact-checking, Transactions of the Association for Computational Linguistics, 10, pp.178-206
Amazeen M. A. (2015), Revisiting the Epistemology of Fact-checking, Critical Review, 27(1), pp.1-22
OECD (2024), Facts not Fakes: Tackling Disinformation, Strengthening Information Integrity, OECD Publishing, Paris, https://doi.org/10.1787/d909ff7a-en

XAI (Explainable AI)

L'Intelligenza artificiale cosiddetta 'spiegabile' è un campo di ricerca che mira a rendere trasparenti e comprensibili le decisioni prese dai modelli di IA complessi. L'obiettivo è passare da sistemi di 'scatola nera' sopra citati a sistemi trasparenti in grado di spiegare il ragionamento dietro le proprie risposte, un requisito fondamentale in settori come la sanità, la finanza e la giustizia, dove la trasparenza è indispensabile.

Fonte: Gunning D. *et al.* (2019), XAI-Explainable Artificial Intelligence, Science Robotics

6. Impatti, governance, etica e regolazione

6.1 Il rapporto tra intelligenza artificiale, diritto, etica e politiche pubbliche

L'intelligenza artificiale non è solo una questione tecnologica: il suo impatto sulla società, sui diritti fondamentali, sul lavoro e sulle dinamiche democratiche richiede un quadro di regole, principi etici e strutture di governance capaci di guidarne lo sviluppo responsabile. Questa sezione raccoglie i termini e i concetti che definiscono il rapporto tra intelligenza artificiale, diritto, etica e politiche pubbliche.

Nella sezione si trova innanzitutto l'AI Act, il Regolamento europeo sull'intelligenza artificiale approvato nel 2024, prima legge al mondo a disciplinare in modo organico l'uso dell'IA attraverso un approccio basato sul rischio. Accanto al quadro normativo europeo, la sezione presenta la Strategia italiana per l'Intelligenza artificiale 2024-2026, il documento che definisce la roadmap nazionale per l'implementazione dell'IA nella ricerca, nella pubblica amministrazione e nel tessuto produttivo del Paese.

Sul piano etico, si scopriranno concetti di AI Governance, che abbraccia l'insieme delle norme e delle strutture organizzative per un uso responsabile dell'IA, e di AI Safety, il campo di ricerca dedicato a garantire che i sistemi intelligenti rimangano sicuri e controllabili. La sezione approfondisce inoltre temi come il bias algoritmico, una delle sfide più pressanti dell'IA responsabile, e l'algoretica, disciplina che esplora il rapporto tra algoritmi e norme morali, nata dal dialogo tra tecnologia, filosofia e teologia e promossa dalla Rome Call for AI Ethics.

Completano la categoria il principio dell'Human-in-the-Loop (HITL), che garantisce la partecipazione umana nei processi decisionali automatizzati soprattutto in ambiti critici, e le riflessioni più ampie sull'etica dell'intelligenza artificiale che coinvolgono questioni di responsabilità, privacy, equità e trasparenza. Questa sezione è fondamentale per chi desidera comprendere come la società, le istituzioni e i legislatori stiano

cercando di governare una tecnologia il cui potenziale trasformativo impone un bilanciamento costante tra innovazione, tutela dei diritti e bene comune.

AI Act

Il Regolamento europeo sull'intelligenza artificiale, approvato dall'Unione Europea nel 2024[14]. È il primo quadro normativo organico e orizzontale adottato a livello sovranazionale per disciplinare i sistemi di IA, classificando i sistemi di IA in base al livello di rischio (da minimo a inaccettabile) e stabilendo regole, obblighi e divieti per sviluppatori e utilizzatori. L'obiettivo è tutelare la salute, la sicurezza e i diritti fondamentali dei cittadini senza frenare l'innovazione.

Fonte: Parlamento europeo e Consiglio dell'Ue. (2024), AI Act: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

AI Governance

L'insieme delle norme, dei principi etici, delle politiche e delle strutture organizzative che guidano lo sviluppo e l'uso responsabile dell'intelligenza artificiale. Comprende aspetti come la trasparenza algoritmica, la protezione dei dati personali, l'accountability dei soggetti che progettano o utilizzano sistemi di decisione automatizzata e la prevenzione dei bias discriminatori.

Fonte: Floridi L. *et al.* (2018), AI4People-An Ethical Framework for a Good AI Society, Minds and Machines Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on Artificial Intelligence (Artificial Intelligence Act), Official Journal of the European Union, L 1689, 12 July 2024

AI Safety

Campo di ricerca che studia come prevenire comportamenti indesiderati e dannosi nei sistemi di *machine learning* derivanti da errori di progettazione, specificazione dell'obiettivo o processo di apprendimento.
In questa prospettiva, un 'incidente' si verifica quando un sistema produce risultati inattesi o nocivi pur avendo ricevuto un obiettivo formalmente definito dal progettista.
L'AI *Safety* si concentra quindi su problemi tecnici concreti, come definire correttamente gli obiettivi del sistema, evitare che il modello "imbrogli" per ottenere risultati (*reward hacking*), garantire un apprendimento sicuro e assicurare che continui a funzionare correttamente anche quando i dati cambiano nel tempo.
In letteratura recente il termine è talvolta utilizzato in senso più ampio per riferirsi anche a problemi di allineamento e rischi sistemici delle IA avanzate.

Fonte: Amodei D. *et al.* (2016), Concrete Problems in AI Safety, arXiv preprint

Algoretica

Campo di riflessione interdisciplinare che analizza le implicazioni etiche degli algoritmi, con particolare attenzione ai criteri di trasparenza, responsabilità, equità e tutela della dignità umana. Il termine è stato introdotto e promosso nel contesto italiano della Rome Call for AI Ethics (2020-2021) per indicare l'esigenza di integrare principi etici nella progettazione e nell'impiego degli algoritmi.

Fonte: Rome Call for AI Ethics (2021), https://www.romecall.org/

[14] Unione europea, Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce norme armonizzate sull'intelligenza artificiale (Artificial Intelligence Act), Gazzetta ufficiale dell'Unione europea, L 1689, 12.7.2024.

Bias algoritmico

Distorsione sistematica nei risultati prodotti da un sistema di intelligenza artificiale, derivante da dati di addestramento non rappresentativi, da scelte di progettazione del modello o da assunzioni implicite incorporate nel processo di sviluppo. Il bias può portare a discriminazioni: ad esempio, un sistema di selezione del personale potrebbe penalizzare candidati di un certo genere o etnia se i dati storici riflettono pregiudizi esistenti. Riconoscere e mitigare i bias è una delle sfide più importanti dell'IA responsabile.

Fonte: Zhang B. H., Lemoine B., Mitchell M. (2018), Mitigating Unwanted Biases with Adversarial Learning, in Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, pp. 335-340
Mehrabi N., Morstatter F., Saxena N., Lerman K., Galstyan A. (2021), A Survey on Bias and Fairness in Machine Learning, ACM Computing Surveys (CSUR), 54(6), pp.1-35

Ethics in AI

L'etica dell'Intelligenza Artificiale è ambito di studio che analizza le implicazioni morali, sociali e normative connesse alla progettazione, allo sviluppo e all'impiego dei sistemi di intelligenza artificiale. Include, tra i suoi nuclei tematici, l'attribuzione della responsabilità per gli output generati o per il supporto ai processi decisionali, la tutela della privacy e dei dati personali utilizzati nei processi di addestramento e di funzionamento, nonché la prevenzione di effetti discriminatori o dell'amplificazione di disuguaglianze sociali nei diversi contesti applicativi. Si configura come campo interdisciplinare che coinvolge filosofia morale e politica, diritto, informatica, scienze sociali e attori della società civile.

Fonte: Floridi L. (2019), AI and Its New Winter, Philosophy & Technology
UNESCO (2021), Recommendation on the Ethics of Artificial Intelligence
Floridi L. (2023), The Ethics of Artificial intelligence: Principles, Challenges, and Opportunities, Oxford University Press

HITL (Human-in-the-Loop)

Si tratta di un approccio in cui l'essere umano partecipa attivamente al processo decisionale dell'intelligenza artificiale, verificando, correggendo o approvando gli output generati dal sistema. È fondamentale in ambiti critici (medicina, giustizia, finanza) dove le conseguenze di errori possono essere gravi, e consente che, in contesti critici, le decisioni finali possano essere soggette a supervisione umana.

Fonte: Holzinger A. (2016), Interactive Machine Learning for Health Informatics, IEEE Journal of Biomedical and Health Informatics

Strategia italiana per l'Intelligenza artificiale 2024-2026

Il documento strategico nazionale che definisce gli obiettivi e le azioni dell'Italia per l'implementazione dell'intelligenza artificiale nella ricerca, nella pubblica amministrazione, nelle imprese e nella formazione. Rappresenta la roadmap del Paese per favorire l'adozione responsabile dell'IA, massimizzando le opportunità e mitigando i rischi attraverso uno sviluppo etico e sicuro.

Fonte: Agid (2024), Strategia italiana per l'Intelligenza artificiale 2024-2026
https://www.agid.gov.it/sites/agid/files/2024-07/Strategia_italiana_per_l_Intelligenza_artificiale_2024-2026.pdf

6.2 Fenomeni, rischi e limiti dell'intelligenza artificiale

Ogni tecnologia potente porta con sé rischi, limiti e fenomeni indesiderati: conoscerli è la condizione essenziale per un utilizzo consapevole e responsabile. Questa sezione chiude la tassonomia affrontando il lato critico dell'intelligenza artificiale, offrendo gli strumenti per riconoscere le insidie, comprendere le vulnerabilità e sviluppare un approccio informato ai sistemi di IA.

In quest'ultima parte della trattazione si descrive il concetto di allucinazione artificiale, uno dei limiti più significativi e conosciuti dell'IA generativa: la tendenza dei modelli a produrre informazioni che appaiono plausibili e convincenti, ma che sono in realtà inventate o errate. A questo si affianca la tendenza dei modelli a fornire risposte compiacenti piuttosto che accurate, confermando le aspettative dell'utente invece di correggerle. Entrambi i fenomeni evidenziano l'importanza del pensiero critico nell'interazione con l'IA.

La sezione esplora poi le minacce alla società e alla sicurezza: i deepfake, contenuti multimediali manipolati dall'IA che rappresentano una seria minaccia per la disinformazione e la privacy; lo slop, l'inondazione di contenuti di bassa qualità generati massicciamente dall'IA che sta degradando la qualità dell'informazione online; e il LLMJacking, una nuova forma di attacco informatico che prende di mira i servizi di IA nel cloud. Accanto a questi, il concetto di algocrazia descrive il rischio di un sistema in cui le decisioni che influenzano la vita delle persone vengono prese da algoritmi senza trasparenza né possibilità di contestazione.

Completano il paragrafo i guardrail, ovvero le barriere di sicurezza integrate nei sistemi di IA, il fenomeno del data drift che erode nel tempo le prestazioni dei modelli, e i concetti speculativi di superintelligenza artificiale (ASI). Questa categoria rappresenta un invito alla consapevolezza: l'intelligenza artificiale è uno strumento straordinario, ma il suo utilizzo informato richiede la comprensione tanto delle sue capacità quanto dei suoi limiti.

Allucinazione artificiale

Fenomeno in cui un modello di IA genera informazioni che appaiono plausibili e convincenti, ma che in realtà sono inventate o errate. Ad esempio, un chatbot potrebbe citare un libro inesistente con titolo e autore credibili, o fornire dati statistici completamente falsi accompagnati da formulazioni linguistiche che li rendono convincenti.
Rappresenta un limite significativo da considerare nell'impiego di modelli di IA generativa, in particolare per applicazioni che richiedono alta affidabilità delle informazioni

Fonte: Google AI Principles: Responsibility in AI Development
Salvagno M., Taccone F. S., Gerli A. G. (2023), Artificial intelligence Hallucinations, Critical Care, v.27(1), https://doi.org/10.1186/s13054-023-04473-y

Zhang, Y., Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., ... Shi, S. (2025), Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models. Computational Linguistics, 51(4), pp.1373-1418

Algocrazia

Un termine che descrive un sistema in cui i processi decisionali che influenzano la vita delle persone sono supportati o fortemente orientati da algoritmi, spesso senza trasparenza né possibilità di contestazione. Dall'accesso al credito alla selezione del personale, dalla sorveglianza urbana ai contenuti mostrati sui social media, gli algoritmi assumono un ruolo sempre più rilevante nel modellare le opportunità e i vincoli che le persone incontrano, a seconda del contesto applicativo e delle regole di governance.

Fonte: https://www.treccani.it/enciclopedia/algocrazia_(Neologismi)
Danaher J. (2016), The Threat of Algocracy: Reality, Resistance and Accommodation, Philosophy & Technology, 29(3), pp.245–268
Flórez Rojas, M. L. (2025), Algocracy in the Judiciary: Challenging Trust in the System, Revista de Estudios Sociales, 93

ASI (Artificial Superintelligence)

La Superintelligenza artificiale è un'ipotesi teorica relativa a un'IA ipotetica capace di eccellere oltre le prestazioni umane in ambiti come il ragionamento complesso, la risoluzione di problemi e la generazione di soluzioni innovative, secondo scenari teorici.
Non è una tecnologia esistente, ma un concetto speculativo che stimola discussioni filosofiche e accademiche sui potenziali rischi e scenari di sicurezza associati a sistemi di IA avanzata.

Fonte: Yampolskiy R. V. (2015), Artificial Superintelligence: a Futuristic Approach, CRC Press

Attack Categories in LLM-Agent

Le seguenti categorie descrivono le principali tipologie di attacco nei sistemi basati su modelli linguistici e agenti.
Gli attacchi di manipolazione dell'input (quali prompt injection, jailbreaking e context hijacking) agiscono attraverso l'interazione linguistica, inducendo il modello a produrre output non conformi. Colpiscono principalmente le componenti con cui l'utente interagisce direttamente, come chatbot, applicazioni e agenti.
Gli attacchi di compromissione del modello (come data poisoning, backdoor e memory injection) incidono sul funzionamento interno del sistema, alterando i dati o i processi di apprendimento.
Gli attacchi a sistema e privacy (tra cui membership inference, retrieval poisoning e side channels) mirano a estrarre informazioni sensibili o a compromettere l'integrità dei dati.
Infine, le vulnerabilità di protocollo (come exploit legati a MCP, ACP, ANP e A2A) riguardano i meccanismi di comunicazione tra componenti e possono compromettere lo scambio di informazioni tra sistemi e agenti.

Fonte: Ferrag M. A., Tihanyi, N., Hamouda D., Maglaras, L., Lakas A., Debbah M. (2025), From Prompt Injections to Protocol Exploits: Threats in LLM-powered AI Agents Workflows

Data Drift

La Deriva dei dati è il fenomeno per cui i dati che un modello di IA incontra nel mondo reale cambiano nel tempo rispetto ai dati su cui è stato addestrato, facendo calare le sue prestazioni.
Ad esempio, un modello addestrato sui comportamenti di acquisto pre-pandemia potrebbe vedere un calo di performance dopo la pandemia, a causa del cambiamento nella distribuzione dei dati dei consumatori. La gestione del data drift richiede quindi attività sistematiche di monitoraggio, validazione e aggiornamento continuo dei modelli, al fine di garantirne l'affidabilità nel tempo.

Fonte: Gama J. *et al.* (2014), A Survey on Concept Drift Adaptation, ACM Computing Surveys

Deepfake

Contenuti multimediali (video, immagini o audio) generati o manipolati dall'intelligenza artificiale per apparire realistici, pur essendo generati o modificati digitalmente e non corrispondenti a eventi reali.
Il termine nasce dalla fusione di *deep learning* e *fake* (falso). I deepfake possono mostrare persone che dicono o fanno cose mai realmente accadute, potenzialmente rilevanti per la disinformazione e la protezione della privacy, a seconda del contesto di utilizzo e delle misure di mitigazione.

Fonte: Parlamento europeo e Consiglio dell'Ue. (2024), AI Act:
https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

Guardrail

I guardrail rappresentano i meccanismi e le regole integrate nei sistemi di IA per limitare la generazione di output dannosi, come contenuti violenti, disinformazione o divulgazione non autorizzata di dati personali, e ridurre il rischio di utilizzo del sistema per attività illecite. Tali barriere di sicurezza servono a evitare che l'IA generi contenuti violenti, diffonda disinformazione, riveli dati personali o assista in attività illecite. Costituiscono un elemento rilevante nelle strategie di sicurezza e mitigazione dei rischi dei sistemi di IA.

Fonte: Parlamento europeo e Consiglio dell'Ue. (2024), AI Act:
https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

Slop

Termine inglese (letteralmente sbobba) diventato popolare dal 2024 per indicare in modo critico i contenuti caratterizzati da ripetitività, limitata originalità e privi di valore generati massicciamente dall'intelligenza artificiale e pubblicati online. Si potrebbe definire una sorta di 'spazzatura digitale' che riempie social e siti web. La diffusione dello slop su social media e siti web è vista come una delle conseguenze negative della democratizzazione degli strumenti di IA generativa. Il dizionario Merriam-Webster lo ha scelto come parola dell'anno per il 2025.

Fonte: de Wynter A., Wang X., Sokolov A., Gu Q., Chen S.-Q. (2023), An Evaluation on Large Language Model Outputs: Discourse and Memorization
Huang Q. (2025), Evaluating the Quality of AI-generated Digital Educational Resources: An Evaluation Index System, Systems, 13(3), 174

Sycophancy

La tendenza di alcuni modelli di IA a rispondere in modo eccessivamente compiacente, generando risposte che tendono a confermare le preferenze percepite dell'utente piuttosto che fornire informazioni accurate o basate su evidenze. È un problema rilevante perché può portare l'utente a ricevere conferme errate delle proprie convinzioni invece di risposte obiettive.

Fonte: Sharma M., Tong M., Korbak T., Duvenaud D., Askell A., Bowman S. R., Perez E. (2023), Towards Understanding Sycophancy in Language Models, arXiv preprint arXiv:2310.13548

Conclusioni

La tassonomia proposta supera una semplice elencazione di concetti e si configura come una struttura analitica che intende spiegare i principi essenziali relativi all'ecosistema dell'IA, chiarire ambiguità terminologiche, collegando aspetti tecnologici, organizzativi e regolatori.

L'obiettivo di voler realizzare analisi empiriche solide, basate su dati affidabili rende necessario poter usufruire di un lessico condiviso e progressivamente organizzato. La classificazione che si propone è innanzitutto una struttura concettuale con un approccio accessibile e non eccessivamente specialistico. Al contempo, essa rappresenta una base di riferimento per progettare indagini, elaborare strumenti di raccolta dati e sviluppare ricerche sui temi del lavoro, delle professioni, della formazione e delle competenze.

Molto spesso, dai questionari somministrati a lavoratori e imprese emerge una evidente incertezza riguardo ai concetti e alle tecnologie più recenti, specie quelli legati all'intelligenza artificiale. Pertanto, per ottimizzare l'affidabilità dei risultati, è utile offrire un sostegno per limitare le ambiguità e ottenere risposte più consapevoli, proponendo definizioni semplici e comprensibili.

L'impianto teorico-analitico del presente modello classificatorio consente di collegare i diversi ambiti dell'IA a specifiche aree di analisi, utili per mappare le competenze e valutare l'impatto delle tecnologie sui processi produttivi e sulle trasformazioni professionali. L'intento è quello di superare i riferimenti generici all'IA, attraverso l'articolazione per funzioni e un orientamento verso categorie operative concrete. Ne consegue una progettazione di strumenti di rilevazione più mirati per imprese e lavoratori resi più consapevoli delle differenti dimensioni operative (modalità d'uso, livelli di integrazione, stadi di maturità tecnologica).

Con l'introduzione dell'IA non si assiste solo alla sostituzione di attività e mansioni, ma a una vera e propria ridefinizione della distribuzione di competenze e responsabilità. Il binomio tecnologie e skill diventa un nodo cruciale su cui questo contributo invita a riflettere. Ad ogni tipologia classificata si possono associare competenze specifiche: interazione con sistemi generativi alla supervisione, verifica degli output, integrazione delle soluzioni nei processi organizzativi. È evidente come questo rinnovato scenario apra la strada a dinamiche di complementarità e di possibile sostituzione tra uomo e macchina. Questo andamento pone la necessità di disporre di strumenti analitici capaci di misurare in modo differenziato l'esposizione delle professioni e l'intensità dei cambiamenti nei compiti.

Le innovazioni introdotte e, dunque, le tecnologie classificate, possono incidere non solo su singole attività, ma su intere fasi dei processi, fino a determinare – nei casi più complessi – una riorganizzazione dei flussi di lavoro. Tali trasformazioni si traducono in forme di supporto graduale, ibridazione operativa o automazione parziale sotto controllo umano, con effetti su tempi, modalità di coordinamento e assetti organizzativi.

La tassonomia offre quindi una chiave di lettura per comprendere come le soluzioni di IA si integrino nei sistemi produttivi e quali effetti producano sulle mansioni e sui percorsi professionali.

Emerge la necessità di promuovere un uso responsabile dell'IA, basato su governance trasparente e valutazioni preventive dei rischi nei settori ad alto impatto sociale. Allo stesso tempo, la tassonomia evidenzia la necessità di rafforzare percorsi di alfabetizzazione all'IA e formazione continua, differenziati per profili professionali, insieme a politiche attive e strategie di *upskilling* e *reskilling* supportate da strumenti di monitoraggio dell'esposizione delle occupazioni.

Ne consegue, per la tassonomia, anche un ruolo orientativo per le politiche pubbliche, offrendo un quadro chiaro per gestire le implicazioni tecnologiche, organizzative ed etico-sociali dell'IA.

In sintesi, la presente tassonomia rappresenta una solida infrastruttura concettuale e metodologica sia per guidare future ricerche sia per supportare la progettazione di politiche pubbliche. Essa fornisce uno

strumento analitico iniziale per valutare l’impatto dell’intelligenza artificiale sul mercato del lavoro, sulle competenze e sull’organizzazione dei processi produttivi, permettendo di interpretare le trasformazioni in corso in maniera coerente e integrata, evitando letture frammentarie.

Bibliografia

Abercrombie G., Benbouzid D., Giudici P., Golpayegani D., Hernandez J., Noro P., Waltersdorfer L. (2024), A collaborative, human-centred taxonomy of ai, algorithmic, and automation harms, arXiv preprint arXiv:2407.01294

Acemoglu D., Restrepo P. (2019), Automation and new tasks: How technology displaces and reinstates labor, Journal of economic perspectives, 33, n.2, pp3-30

Autor D., Salomons A. (2018), Is automation labor-displacing? Productivity Growth, Employment, and the Labor Share, NBER Working Papers, 24871, <https://ideas.repec.org/p/nbr/nberwo/24871.html>

Bresnahan T. F., Trajtenberg M. (1995), General Purpose Technologies Engines of Growth?, Journal of econometrics, 65, n.1, pp.83-108

Brynjolfsson E., Li D., Raymond L. (2025), Generative AI at work, The Quarterly Journal of Economics, 140, n.2, pp. 889-942

Cherner T., Egan A., Howe K., Whitfield E. J. (2025), An Essential Glossary for Generative Artificial Intelligence in Higher Education, AI Enhanced Learning, 1, n.1, pp.61-98

Commissione europea (2019), Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the regions, Building Trust in Human-Centric Artificial Intelligence, Brussels, 8.4.2019 COM (2019) 168 final

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

Estevez Almenzar M., Fernandez Llorca D., Gomez E., Martinez Plumed, F. (2022), Glossary of Human-Centric Artificial Intelligence, Publications Office of the European Union, Luxembourg, https://publications.jrc.ec.europa.eu/repository/handle/JRC129614

EIT (2021), Creation of a taxonomy for the European AI ecosystem: A report of the Cross-KIC activity "Innovation Impact Artificial Intelligence", European Institute of Innovation & Technology, Brussels, https://www.eit.europa.eu/sites/default/files/creation_of_a_taxonomy_for_the_european_ai_ecosyste m_final.pdf

Ferri V., Porcelli R., Pelucchi M. (2025), Le conoscenze tecnico-specialistiche in materia di IA: un'analisi degli annunci di lavoro online del 2024, Roma, Inapp, Focus Inapp n. 14

Ferri V., Porcelli R., Fenoaltea E. M. (2024), Lavoro e Intelligenza artificiale in Italia: tra opportunità e rischio di sostituzione, Roma, Inapp, WP, 125

Galindo-Cuesta J. A. (2025), Glossary of Generative Artificial Intelligence for Education: A Conceptual and Pedagogical Framework, Review of Artificial Intelligence in Education, n.6 <https://educationai-review.org/revista/article/view/47/40>

Kundisch D., Muntermann J., Oberlaender A., Rau D., Röglinger M., Schoormann T., Szopinski D. (2022), An Update for Taxonomy Designers, Business & Information Systems Engineering, 64, n.4, pp.421-439

Marsiglia S., Ferri V., Fiore A., Tesauro G. (a cura di) (2025), Esposizione delle professioni all'intelligenza artificiale e indicazioni di policy, Arti-Inapp, <https://oa.inapp.gov.it/handle/20.500.12916/5203>

Newman, J. (2023), A taxonomy of trustworthiness for artificial intelligence, UC Berkeley Center for Long-Term Cybersecurity, North Charleston, SC

OECD (2024), A sectoral taxonomy of AI intensity - OECD Artificial Intelligence, Papers December 2024 n.30

Paeplow J., Schoormann T., Möller, F., Strobel G. (2025), AI Startups for Good: A Taxonomy and Archetypes of Sustainable Business Models, Journal of Cleaner Production, n.520, <https://www.sciencedirect.com/science/article/pii/S0959652625014945>

Strobel G., Banh L., Möller, F., Schoormann T. (2024), Exploring Generative Artificial Intelligence: A Taxonomy and Types, Hawaii International Conference on System Sciences (HICSS)

Theofanos M., Choong Y., Jensen T. (2024), AI Use Taxonomy: A Human-Centered Approach, NIST Trustworthy and Responsible AI, National Institute of Standards and Technology, Gaithersburg, MD, March 2024

Williams B. A. (2025), The Influence of Perceived Anthropomorphism and Social Presence on AI Interface User Experience: A Systematic Review, International Journal of Human–Computer Interaction, pp.1-18

INAPP
ISTITUTO NAZIONALE PER L'ANALISI
DELLE POLITICHE PUBBLICHE

COLLANA
FOCUS
Inapp