Determinare il mismatch delle competenze usando la GenAI

Costruzione di un modello su istruzione, formazione e mercato del lavoro in Italia

Alessia Romito
INAPP

Boris Sofronic
INAPP

Lo studio analizza l'utilizzo dei *Large Language Models* (LLM) per l'individuazione del mismatch delle competenze in Italia. Confrontando due sperimentazioni condotte nel 2023 e nel 2025, evidenzia l'evoluzione dei modelli di Intelligenza artificiale generativa, dai limiti iniziali (errori fattuali, obsolescenza della knowledge-base, allucinazioni) ai recenti progressi (Chain-of-Thought, RLHF, RAG). Il lavoro identifica metodologie efficaci per l'impiego rigoroso dei LLM nella ricerca: selezione accurata delle fonti, personalizzazione mirata, prompt-engineering strategico e approccio ibrido con supervisione umana.

This study analyses the use of Large Language Models (LLMs) to identify skills mismatch in Italy. By comparing two experiments conducted in 2023 and 2025, it traces the evolution of generative AI models, from early limitations (factual inaccuracies, obsolescence of knowledge bases, hallucinations) to recent advancements such as Chain-of-Thought prompting, Reinforcement Learning from Human Feedback (RLHF), and Retrieval-Augmented Generation (RAG). The study identifies effective methodologies for rigorous LLM application in research: accurate source selection, targeted customisation, strategic prompt engineering, and a hybrid approach involving human oversight.

DOI: 10.53223/Sinappsi_2025-02-14

Citazione
Romito A., Sofronic B. (2025), Determinare il mismatch delle competenze usando la GenAI. Costruzione di un modello su istruzione, formazione e mercato del lavoro in Italia, *Sinappsi*, XV, n.2, pp.208-227

Parole chiave
Intelligenza artificiale
Mercato del lavoro
Skill mismatch

Keywords
Artificial intelligence,
Labour market
Skills mismatch

Introduzione[1]

I modelli linguistici di grandi dimensioni (*large language models*, LLM) sono modelli di Intelligenza artificiale (IA) generativa (GenAI) che utilizzano l'apprendimento automatico (*machine learning*, ML) per comprendere e generare linguaggio umano. Sono basati su reti neurali artificiali (*artificial neural networks*, ANN) e tecniche di elaborazione del linguaggio naturale (*natural language processing*, NLP). A marzo 2025 i LLM più conosciuti e più usati sono

1 Alla fine del documento è presente il glossario dei termini più tecnici.

i *Generative Pre-trained Transformer* (GPT) come: ChatGPT della OpenAI, Microsoft Copilot, Claude.ai della Antrophic, DeepSeek, Google Gemini, Llama della Meta, Grok della xAI, Mistral AI e molti altri. Questi modelli si basano su un'architettura chiamata *Transformer*, un'innovazione fondamentale nel campo delle ANN, introdotta da ricercatori di Google Brain e Google Research nel paper *Attention Is All You Need* (Vaswani *et al.* 2017). I Transformer sono progettati per analizzare le relazioni e le dipendenze tra gli elementi di una sequenza, come le parole in una frase, utilizzando meccanismi di *Self-attention* per determinare l'importanza di ogni elemento e generare risposte sempre più accurate (*ibidem*, 3).

Le sperimentazioni sui LLM da parte della comunità scientifica internazionale hanno prodotto risultati misti. Nonostante gli entusiasmanti progressi offerti dai LLM, la letteratura scientifica documenta anche limiti significativi; ad esempio, sono stati osservati errori fattuali (Augenstein *et al*. 2023), allucinazioni artificiali (Alkaissi e McFarlane 2023; Wang *et al.* 2023), errori causati dai dati reperiti da fonti esterne o errori di inferenza (Wang *et al.* 2023, 5)[2]. Questo testo descrive alcune esperienze di utilizzo sperimentale dei LLM nell'individuazione del mismatch delle competenze e, più in generale, l'uso dei LLM nei domini dell'istruzione, della formazione, delle professioni e dell'occupazione condotte in Inapp e ha come obiettivo l'identificazione delle metodologie per rilevare errori fattuali e allucinazioni, analizzarne le cause e sviluppare strategie per un uso consapevole dell'IA nella ricerca scientifica. Il metodo è stato già sperimentato da Chen *et al.* (2024) seppur con finalità differenti.

Come vedremo di seguito, l'IA consiste di un ampio panorama di tecnologie progettate per consentire alle macchine di percepire, interpretare, apprendere e agire emulando le capacità cognitive umane. All'interno di questa gamma delle soluzioni IA, i LLM e la GenAI occupano un posto centrale, essendo in grado di creare contenuti originali, dai testi alle immagini, fino ad audio e video. Altri modelli IA, invece, si concentrano su compiti specifici, come il riconoscimento di pattern, ottimizzando processi ripetitivi per aumentare la produttività piuttosto che creare nuovi contenuti (Cazzaniga *et al*. 2024). In attesa della cosiddetta Intelligenza artificiale generale (AGI), tecnologie più tradizionali e mature, come i sistemi di ML basati sulle librerie open source come PyTorch (Ansel *et al*. 2024) e TensorFlow[3] o piattaforme commerciali come IBM Watson Studio, Google Cloud AutoML e Microsoft Azure Machine Learning Studio, continuano a dimostrarsi più utili per assistere la ricerca scientifica (Singh 2023). Questo è particolarmente vero nei settori dell'istruzione e dell'occupazione, dove la qualità dei dati e l'accuratezza dell'analisi sono fondamentali.

1. I confini regolamentari e istituzionali della sperimentazione

Le tecnologie basate sull'Intelligenza artificiale stanno diventando sempre più pervasive, mostrando il loro potenziale trasformativo sulle dinamiche sociali, quali ad esempio la formazione, l'occupazione, la ricerca, nonché su quelle produttive. A fronte di una indispensabile necessità di accogliere l'innovazione, è emerso un altrettanto imprescindibile bisogno di regolamentarne l'uso, allo scopo di garantire che i sistemi di IA siano sicuri, etici e affidabili e che siano sviluppati e utilizzati in modo responsabile.

Il Consiglio europeo ha fatto un passo decisivo in questa direzione adottando il *Regolamento europeo sull'Intelligenza Artificiale*[4] (AI Act), entrato in vigore il 1° agosto 2024. Il documento affronta i potenziali rischi per la salute, la sicurezza e i diritti fondamentali dei cittadini, fornisce requisiti e obblighi a sviluppatori e operatori per quanto riguarda gli usi

2 Si veda la figura 1 del presente articolo.

3 TensorFlow Developers, *TensorFlow: Large-scale machine learning on heterogeneous systems* (Computer software), https://doi.org/10.5281/zenodo.4724125.

4 Parlamento europeo, Consiglio dell'Unione europea (2024), *Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (Regolamento sull'Intelligenza artificiale)*, https://eur-lex.europa.eu/eli/reg/2024/1689/oj.

specifici dell'IA e introduce un quadro uniforme in tutti i Paesi dell'UE basato sul rischio[5].

A livello nazionale, un primo modello di governance dell'IA in Italia viene elaborato a maggio del 2024, quando il Dipartimento per il Programma del Governo della Presidenza del Consiglio dei ministri pubblica un focus in materia di IA che analizza l'AI Act e presenta un Disegno di Legge (DDL) per l'approvazione parlamentare[6]; nel medesimo periodo il Dipartimento per la Trasformazione digitale e l'Agenzia per l'Italia digitale (Agid) pubblicano il documento integrale della *Strategia Italiana per l'Intelligenza artificiale 2024-2026* (Agid 2024). Il documento pone particolare attenzione all'implementazione dell'IA nella ricerca, nella pubblica amministrazione (PA), nelle imprese e nella formazione, definendo obiettivi e azioni strategiche per ciascun ambito. Per quanto riguarda la ricerca scientifica, l'Agid suggerisce che l'Italia dovrà 'rafforzare gli investimenti nell'IA, promuovere la creazione di competenze di ricerca e tecnologie calate nel contesto del Sistema-Paese in linea con principi di affidabilità e responsabilità, propri ai valori antropocentrici dell'UÈ, rimarcando la 'libertà di sperimentazione, utilizzando contenuti e dati per la creazione di dataset e l'addestramento di modelli resi disponibili in open source' (*ibidem*, 9).

Tra le strategie contemplate nella macroarea ricerca, coerenti con le finalità dello studio presentato, si segnalano: "il consolidamento dell'ecosistema italiano della ricerca, la progettazione di LLM italiani, progetti interdisciplinari per il benessere sociale, la ricerca fondazionale e blu-sky per l'IA di prossima generazione" (*ibidem*, 14).

È all'interno di questo perimetro, delineato da un quadro giuridico, europeo e nazionale, e di indirizzo, che l'Inapp è stato chiamato ad agire da parte del Ministero del Lavoro[7]. Nel documento programmatico *Indirizzi strategici triennio 2025-2027*[8], tra le tematiche di rilevanza strategica, viene promossa "l'analisi e la valutazione degli impatti delle tecnologie digitali in generale (e più nello specifico dell'Intelligenza artificiale) sull'occupazione, sui fabbisogni di aggiornamento e di riconversione delle competenze […]". In questo contesto si inseriscono le sperimentazioni sull'IA già svolte nell'ambito delle attività dell'Istituto con lo scopo di fornire spunti utili a creare una base di partenza per le attività future che l'ente intenderà avviare in questa direzione e condividere con la comunità di ricerca informazioni sui risultati raggiunti e criticità riscontrate. Un'esperienza senz'altro interessante è stato il progetto Sophia in cui l'Inapp ha sperimentato l'implementazione di diverse soluzioni IA in una piattaforma di Labour Market Intelligence (LMI), con l'intento di misurare e segnalare il mismatch delle competenze analizzando le *web job vacancies*, risultati di apprendimento delle opportunità formative e CV degli utenti (Sofronic 2022). Il progetto mirava a interconnettere le fonti dati sulle opportunità di apprendimento e quelle sulla domanda del mercato del lavoro usando le tecniche di ML, NLP e gli algoritmi di IA che riescono a collegare direttamente i concetti estratti dai testi destrutturati, ad es. Named Entity Recogniton (NER) o l'IA cognitiva. L'output desiderato era un prototipo funzionante di un LMI in grado di rispondere alle domande sulle competenze, professioni e *soft skills* più richieste, i settori economici più dinamici, le qualificazioni più spendibili per la crescita professionale, le tendenze a medio termine e molto altro. Prima di concentrarsi sulle successive fasi di sviluppo del progetto, è stato necessario e doveroso effettuare un'analisi preliminare accurata, mirata a verificare empiricamente l'affidabilità dei responsi generati dai modelli IA allora disponibili, rispettando i criteri del rigore scientifico.

5 Il Regolamento stabilisce quattro livelli di rischio e ne definisce specifiche regole (https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai):
- rischio minimo: i filtri spam e i videogiochi che sfruttano l'AI non sono soggetti ad alcun obbligo ai sensi del Regolamento;
- rischio specifico per la trasparenza: i sistemi come le chatbot devono informare chiaramente gli utenti che stanno interagendo con una macchina, mentre alcuni contenuti generati dall'IA devono essere etichettati come tali;
- rischio alto: i software medici basati sull'AI o i sistemi di AI utilizzati per il reclutamento, devono rispettare requisiti rigorosi, tra cui sistemi di mitigazione del rischio, set di dati di alta qualità, informazioni chiare sugli utenti, supervisione umana ecc;
- rischio inaccettabile: i sistemi di AI che permettono l'attribuzione di un 'punteggio sociale' da parte di governi o imprese sono considerati una chiara minaccia per i diritti fondamentali delle persone e sono pertanto vietati.

6 Atto Senato n. 1146 si veda https://www.senato.it/leggi-e-documenti/disegni-di-legge/scheda-ddl?did=58262.

7 Cfr. *Indirizzi strategici triennio 2025-2027*, Inapp (Delibera CdA 25 novembre 2024, n. 18).

8 Inapp, Delibera CdA 25 novembre 2024, n. 18, cit.

2. I risultati della prima sperimentazione

Il 30 novembre del 2022 entra in produzione la prima versione pubblica di ChatGPT che, con cento milioni di utenti attivi al mese, dopo un solo bimestre dal lancio, si impone come l'applicazione in più rapida crescita della storia (Augenstein *et al*. 2023). Grazie alle recensioni all'epoca molto positive da parte degli esperti del settore, alla facilità di accesso e alla possibilità di utilizzo limitato gratuito, da gennaio del 2023 il ChatBot di OpenAI si afferma come principale piattaforma di verifica empirica della qualità delle risposte generate da un LLM. Tali considerazioni hanno incoraggiato la sperimentazione degli strumenti di GenAI anche a supporto del modello di Sophia, già basato sulle tecnologie di IA mature e affermate, come ML e NLP.

Nel box 1[9] si riporta una prima interazione con il ChatGPT mirata a verificare in che misura la sua epistemologia fosse aggiornata alle nuove competenze richieste, alle professioni emergenti e alla disponibilità delle rispettive opportunità di apprendimento utili al reskilling e upskilling.

I risultati di questa prima conversazione sembravano incoraggianti e promettenti. Le risposte generate da ChatGPT risultavano logiche, coerenti e ben elaborate. Una breve verifica a campione ha confermato l'esistenza all'epoca di alcuni corsi di laurea, master e dottorato menzionati dal modello.

Le successive interazioni con il modello hanno prodotto risultati misti. Nelle risposte fornite sono stati riscontrati alcuni errori, talvolta critici. Alcuni autori li definiscono come *errori fattuali* (Augenstein *et al.* 2023), alcuni come *allucinazioni artificiali* (Alkaissi e McFarlane 2023), altri invece hanno preferito approfondire l'argomento e distinguere gli errori di risposta in diverse categorie con lo scopo di individuare meglio le cause e proporre eventuali rimedi (Wang *et al*. 2023). Nel paper *Survey on Factuality in Large language models: knowledge, retrieval and domain-specificity* (*ibidem*, 5), gli autori definiscono tutti gli errori riscontrati come fattuali, ma propongono una distinzione tra errori causati al livello del LLM, errori generati dai dati reperiti da fonti esterne ed errori di inferenza (figura 1).

Risulta particolarmente utile concentrare l'attenzione sulla prima e sulla terza categoria (gli errori causati al livello del modello e gli errori causati dall'inferenza), in quanto in buona parte corrispondono all'esperienza riscontrata durante le sperimentazioni all'epoca condotte in Inapp. La ricerca ha, infatti, rilevato tre tipologie principali di errori nelle risposte del

Figura 1. Esempi di diversi tipi di errori fattuali generati dai LLM

Category	Cause	Example Dialog	Notes and references
Model-level causes	Domain knowledge deficit	**Q:** CEO of Assicurazioni Generali? **BloombergGPT:** Philippe Donnet **GPT-NeoX:** Antonio De Lorenzo, Simone Gambarini, Enrico Zanetti **FLAN-T5-XXL:** John M Forsyth, Christopher K Peters, {empty string}	BloombergGPT is a finance domain-specific language model. [278]
	Outdated information	**Q:** When was Kyiv attacked by Russia? **ChatGPT:**As of my last knowledge update in September 2021, Russia had not launched an attack on Kyiv.	Kyiv was attacked by Russia on 25 February 2022.
	Reasoning error	**Q:** Who is Tom Cruise's mother? **A:** Mary Lee Pfeiffer **Q:** Who is Mary Lee Pfeiffer's son? **A:** There is no widely known information about...	From Berglund et al. [13]. It is clear that the model knows Tom Cruise's mother is Lee Pfeiffer, but it fails to reason that Lee Pfeiffer has a son named Tom Cruise.
Inference-level causes	Snowballing	**Q:** Was there ever a US senator who represented the state of New Hampshire and whose alma mater was the University of Pennsylvania? **A:** Yes... His name was John P. Hale	John P. Hale was graduated from Bowdoin College. [301]
	Exposure bias	**Q:** Aamari was very stoic. [PRONOUN] rarely showed any emotion. **A:** He.	The correct answer was Xe according to [96].

Fonte: Wang *et al*. (2023)

9 I box (da 1 a 9) sono consultabili a fine articolo.

modello: 1) scarsa conoscenza del dominio, 2) obsolescenza della knowledge-base, 3) allucinazioni parziali costruite usando elementi veri (errori di inferenza), di seguito esaminati.

La scarsa conoscenza del dominio è stata riscontrata in una delle interazioni con il ChatGPT – di cui un estratto è riportato nel box 2 – che aveva come obiettivo la verifica della conoscenza dei principali sistemi classificatori necessari per individuare meglio il mismatch delle competenze nei diversi settori economici e le professioni.

Il ChatBot ha generato una risposta errata rispetto alla conoscenza della classificazione Istat CP2011, fondamentale per le finalità della sperimentazione. Pur ammettendo di aver fornito un'informazione distorta e accettando la 'correzione', il modello risultava quindi essere poco accurato e, di conseguenza, poco affidabile ai fini della ricerca.

Considerando tuttavia l'evoluzione tecnologica, il continuo apprendimento del ChatBot e in particolar modo facendo affidamento al prompt-engineering[10], a distanza di circa un anno e mezzo, si è ritenuto doveroso effettuare la medesima interrogazione, allo scopo di verificare se il modello fosse in grado di generare un responso corretto alla stessa domanda. L'interrogazione ha prodotto un esito alquanto positivo, la risposta generata infatti mostra elementi di maggior accuratezza (box 3) e conduce a ritenere plausibile che l'interattività tra il modello e l'essere umano può contribuire all'evoluzione del LLM.

Rispetto alla obsolescenza della knowledge-base, si riporta una conversazione con ChatGPT relativa alle eventuali opportunità formative, presenti nel territorio della Provincia di Roma, per acquisire la qualificazione di estetista (box 4), il cui profilo è regolamentato nel Repertorio nazionale dell'istruzione e della formazione professionale (Operatore del benessere - Erogazione dei servizi di trattamento estetico).

Lo scopo della conversazione era quello di verificare se l'epistemologia del modello interrogato permetteva l'interconnessione tra i dati sui centri di formazione accreditati (CFP) censiti da Inapp nel Portale Accreditamento, le qualificazioni contenute nell'Atlante del Lavoro e delle qualificazioni e le rispettive opportunità formative disponibili per le iscrizioni ai corsi da erogare nel prossimo futuro utili a conseguire tali qualificazioni se presenti nei siti web dei singoli enti. La risposta generata dal LLM presentava informazioni errate. Rispetto alle allucinazioni artificiali, occorre fare riferimento alla fase di realizzazione di una delle azioni previste nel progetto Sophia, nello specifico l'attività di analisi delle piattaforme informatiche per l'istruzione e la formazione professionale in Europa e lo studio per una loro interoperabilità.

Si è ritenuto interessante 'mettere alla prova' il LLM, la cui sperimentazione sull'accuratezza proseguiva in parallelo, per individuare le fonti scientifiche e accademiche sui temi d'interesse e in linea con i contenuti trattati, interrogando il modello sulla richiesta di riferimenti bibliografici relativi all'argomento ICT tools on E&T in Europe. Gli esiti dell'interrogazione, riportati nel box 5, sono stati verificati attraverso il motore di ricerca dedicato (Google Scholar) e attraverso specifici strumenti di ricerca di documenti di natura accademica e scientifica; il risultato ottenuto dal confronto ha evidenziato il fenomeno delle allucinazioni artificiali poiché nessun riferimento bibliografico ha trovato corrispondenza in una fonte reale.

La prima sperimentazione sull'affidabilità e l'accuratezza dei LLM, condotta prevalentemente nella prima parte del 2023 usando ChatGPT – unico strumento di GenAI disponibile all'epoca – ha evidenziato diverse criticità (scarsa conoscenza del dominio, obsolescenza della *knowledge-base* e allucinazioni parziali) che meritano un'attenta riflessione. Le risposte generate da ChatGPT riguardo a sistemi classificatori come la CP2011 hanno mostrato una comprensione inizialmente errata del dominio, sebbene successivamente corretta. Questo miglioramento tra interrogazioni distanziate nel tempo suggerisce un possibile aggiornamento della knowledge-base, ma non garantisce l'assenza di errori in altri ambiti specialistici.

La problematica dell'obsolescenza delle informazioni, evidenziata nel caso dell'ente di formazione non più attivo, rappresenta un limite significativo per l'utilizzo di questi strumenti in contesti dinamici come quello delle opportunità formative e del mercato del lavoro. Il fenomeno delle 'allucinazioni artificiali' si è manifestato in modo particolarmente evidente nel caso dei riferimenti bibliografici errati,

10 L'arte di formulare domande a un LLM in modo efficiente per ottimizzare le risposte in funzione delle nostre esigenze (Bolognesi 2024).

che non sono risultati rintracciabili neppure attraverso strumenti di ricerca specialistici. Questo aspetto solleva interrogativi fondamentali sull'affidabilità dei LLM in contesti che richiedono precisione e verificabilità delle fonti, come quello della ricerca scientifica.

Nonostante queste criticità, va riconosciuto che i LLM offrono potenzialità considerevoli come strumenti di supporto all'analisi e all'elaborazione di informazioni complesse. Tuttavia, la sperimentazione ha messo in luce la necessità di adottare un approccio critico e consapevole, che preveda sempre la verifica delle informazioni ottenute attraverso fonti attendibili.

In conclusione, i risultati della prima sperimentazione hanno suggerito che l'utilizzo dei LLM nell'ambito dell'analisi del mismatch richiede cautela e opportuni sistemi di validazione. Il loro impiego ottimale si configura come strumento complementare all'expertise umana, non come sostituto della verifica empirica e dell'analisi critica. Al fine di scongiurare errori fattuali critici e analisi distorte, si è ritenuto necessario sospendere momentaneamente la ricerca sull'utilizzo della GenAI come strumento per l'individuazione del mismatch delle competenze.

3. I risultati della seconda sperimentazione

A distanza di due anni dalla prima sperimentazione, la ricerca è stata ripetuta, questa volta concentrandosi sulle metodologie ibride che integrino le potenzialità dei LLM con sistemi di verifica dell'accuratezza basati su banche dati certificate e aggiornate, come quelle gestite da enti istituzionali, usando i LLM di ultima generazione potenziati dalle nuove funzionalità offerte da modelli 'di frontiera' come le ultime release di ChatGPT 4o-mini, Claude 3.7 Sonnet, DeepSeek V3-R1, Grok 3, Gemini.

In generale, i principali progressi dei modelli GenAI/LLM dal 2023 a marzo 2025 includono:

- tecniche che guidano il modello a 'ragionare' passo-passo, riducendo errori logici, come Chain-of-Thought (Ton *et al.* 2024);
- dataset di pre-addestramento più puliti con rimozione di contenuti tossici o inaffidabili;
- reinforcement learning (Sutton e Barto 2018) avanzato (RLHF);
- integrazione con motori di ricerca in tempo reale per ridurre dipendenza da dati statici (RAG);
- capacità computazionali potenziate da GPU di nuova generazione;
- tecniche di quantizzazione (es. AWQ) per eseguire modelli complessi su hardware limitato.

Di conseguenza, i benchmark come TruthfulQA (Lin *et al.* 2021) mostrano una riduzione delle allucinazioni del 15-30% per modelli top, anche se i migliori modelli mostrano tassi di allucinazione del 5-10% in contesti complessi (es. medicina legale).

In sintesi, la riduzione delle allucinazioni è frutto di progressi sistemici sui dati di training e l'architettura dei modelli, con RAG e RLHF come driver principali. Tuttavia, la complessità del linguaggio naturale mantiene il problema parzialmente irrisolto, richiedendo approcci ibridi uomo-IA per casi critici. Come primo test della seconda sperimentazione, ai principali LLM del mercato sono state poste le stesse domande di quella precedente; i risultati ottenuti sono stati nettamente migliori. Si è evidenziato, infatti, come ChatGPT Plus con le funzioni 'Cerca' e 'Avvia il ragionamento' attivate, riesce nel compito di fornire riferimenti bibliografici, fornendo DOI e URL verificabili dei testi accademici e istituzionali indicati nelle sue risposte (box 6).

Similmente, molto più accurate risultano essere oggi le risposte dei LLM di ultima generazione ai quesiti sulla conoscenza delle classificazioni relative ai domini d'interesse (es. CP2021, ISCED ecc.) e ai quesiti che richiedono una knowledge-base sempre aggiornata (es. centri di formazione professionale con corsi attivi che permettono di acquisire una certa qualificazione/competenza richiesta dal mercato del lavoro). Non tutti i LLM hanno dimostrato, però, lo stesso livello di qualità delle risposte a tutte le domande poste durante il test, in ogni caso resta inteso che lo scopo di questo studio è l'uso della GenAI nella ricerca scientifica come strumento per l'individuazione del mismatch delle competenze in generale, senza alcuna intenzione di benchmarking dei singoli modelli.

L'individuazione e l'addestramento del modello

Tra diversi modelli di GenAI disponibili nel 2025 sul mercato, la scelta del LLM da utilizzare per la seconda sperimentazione è ricaduta sull'ultima release di Anthropic AI - Claude 3.7 Sonnet, versione 'Pro'. I modelli presi in considerazione sono stati valutati usando i dati più recenti dei benchmark maggiormente in uso nell'ambito della ricerca, la data dell'ultimo addestramento/aggiornamento,

presenza/assenza delle funzioni avanzate che permettono un addestramento aggiuntivo sui dati propri e altre personalizzazioni definite dall'utente[11], il rapporto prezzo/prestazioni di tali funzioni avanzate e, infine, la diffusione e la facilità di accesso del modello per permettere la verifica e ripetibilità della sperimentazione da parte della comunità scientifica. Rispetto a questi parametri si è proceduto ad una classificazione che ha condotto a individuare tre modelli di LLM, ai quali è stato somministrato un set di domande (box 7) volte a determinare la loro capacità di individuare – in ottica previsiva – informazioni puntuali e coerenti relative al mismatch in Italia.

I risultati ottenuti – opportunamente editati, omologati nel layout e anonimizzati – sono stati sottoposti a una valutazione di merito in modalità blind, sulla base di criteri di correttezza ed esaustività delle informazioni restituite e assenza di allucinazioni artificiali. In esito all'analisi effettuata la scelta per la sperimentazione è ricaduta sul LLM Claude 3.7 Sonnet[12], che ha presentato il più alto livello di rating rispetto agli indicatori adottati. Una volta creato il progetto 'Determinare il mismatch usando l'Intelligenza artificiale generativa', il nostro modello personalizzato è stato istruito sul suo ruolo, scopo, limitazioni, target e metodologia di ragionamento (box 8).

Successivamente il modello è stato *post*-addestrato prevalentemente con dati e documenti da fonti istituzionali e con ulteriore reportistica che, pur non avendo una rappresentatività statistica, arricchisce il modello con informazioni di dettaglio richieste dal progetto (box 9).

Valutazione dell'accuratezza. Verifica fattuale

Il modello è stato di seguito sollecitato con un quesito specifico e più dettagliato nella sua formulazione, volto a richiedere gli indirizzi di studio dei percorsi ITS e i percorsi universitari più richiesti dal mercato del lavoro e le competenze a essi associate. È stato, quindi, invitato a utilizzare solo fonti statisticamente rilevanti e a omettere qualsiasi analisi di contesto, panoramica e/o descrizione dei percorsi e altre considerazioni. È stato, inoltre, richiesto di restituire le informazioni per topic (indirizzi di studio e competenze associate), indicando per ciascun ambito la fonte informativa, anche se contraddittoria. Le risposte del LLM sono risultate puntuali e coerenti con la richiesta per gli indirizzi di studio dei percorsi ITS, diversamente si è manifestata una distorsione informativa (*bias*) rispetto ai percorsi universitari. In questo caso il modello ha dato autonomamente priorità alla fonte "più ricca di dati di dettaglio" (Rapporti AlmaLaurea su profilo e condizione occupazionale dei laureati), seppur non afferente a dati statisticamente rilevanti, come indicato nell'interrogazione.

Più complesso è il risultato sulle competenze associate. Rispetto agli indirizzi ITS, in assenza di dati nei file di training e nel proprio knowledge-base, il modello ha generato risposte non corrispondenti alla realtà, associandole tra l'altro – in maniera puramente casuale – alle fonti informative da cui ha estratto i dati del primo quesito, generando così allucinazioni artificiali. Tutt'altro risultato si è manifestato, invece, per le competenze associate ai percorsi universitari; il LLM, infatti, ha restituito risposte veritiere, sfruttando sia fonti statisticamente rilevanti, sia altri file di training, presumibilmente la relazione finale del progetto Sophia, ossia l'analisi dei dati estratti dagli annunci di lavoro – di circa un milione di inserzioni sulla piattaforma Monster in Italia – pubblicati nel periodo gennaio-dicembre 2024.

Il modello è stato ulteriormente stimolato con un successivo quesito relativo alle previsioni occupazionali (2-5 anni) per settore economico; in questo caso il LLM ha restituito risposte corrette ed esaustive, beneficiando di fonti attendibili, file di training e informazioni del suo knowledge-base.

Le sperimentazioni illustrate hanno permesso di focalizzare elementi sostanziali e imprescindibili per la creazione di un modello di GenAI efficace nell'individuare il mismatch. Risulta difatti importante, nella fase di training, ricorrere esclusivamente a banche dati attendibili per scongiurare la restituzione di informazioni non coerenti; fare attenzione in fase

11 Ad esempio "DeepThink" e "Search" del DeepSeek V3-R1, "Extended Thinking Mode" di Claude 3.7, "Cerca" e "Avvia il ragionamento" del ChatGPT.

12 Claude 3.7 Sonnet, nella sua versione 'Pro', offre molteplici possibilità di personalizzazione: creazione dei progetti 'custom'; scelta dello stile di risposte (normale, concisa, esplicativa, formale); possibilità di creare stili di risposta personalizzati impostando tono, voce, vocabolario, livelli di dettaglio e molto altro ancora; creazione della propria 'Project knowledge' comprensiva della definizione delle istruzioni al modello e la possibilità di caricare dati aggiuntivi da analizzare prima di fornire le risposte.

Tabella 1. Quadro sinottico comparativo degli indirizzi di studio dei percorsi ITS. Analisi fattuale

Fonte	Interrogazione	Fattuale (*Fonte:* Excelsior-Unioncamere)
Fonte: Indire, XI Monitoraggio nazionale 2024 sui percorsi ITS Academy	1. Tecnologie dell'informazione e della comunicazione: tasso di occupazione dell'89,3% 2. Sistema meccanica: tasso di occupazione dell'87,8% 3. Mobilità sostenibile: tasso di occupazione dell'87,6% 4. Tecnologie innovative per i beni e le attività culturali-turismo: tasso di occupazione dell'85,9% 5. Nuove tecnologie per il made in Italy: tasso di occupazione dell'85,1% 6. Efficienza energetica: tasso di occupazione dell'81,7%	1. Sistema meccanica (38,2%) 2. Metodi e tecnologie per lo sviluppo di sistemi software (13,9%) 3. Architettura e infrastrutture per i sistemi di comunicazione (11,2%) 4. Turismo e attività culturali (7,6%) 5. Processo e impianti a elevata efficienza e a risparmio energetico (7,1%) 6. Servizi alle imprese (7,0%) 7. Sistema moda (5,2%) 8. Mobilità delle persone e delle merci (5,0%) 9. Organizzazione e fruizione dell'informazione e della conoscenza (4,8%)
Fonte: MUR, Dati sugli esiti occupazionali ITS Academy 2023	Sistema meccanica e meccatronica ICT - Information and Communication Technology Mobilità sostenibile e logistica avanzata Automazione industriale Nuove tecnologie per il Made in Italy - settore manifatturiero	

Fonte: elaborazione degli Autori

di prompt-engineering a formulare quesiti il più possibile dettagliati e curati nel lessico, fornendo informazioni guida al LLM utili al recupero delle informazioni accurate.

I risultati ottenuti dai test effettuati, relativamente agli indirizzi di studio dei percorsi ITS, sono stati raccolti in un quadro sinottico (tabella 1) e sono stati confrontati, in ottica fattuale, con una ulteriore fonte di dati di natura istituzionale, espressamente dedicata a monitorare le prospettive occupazionali e i fabbisogni professionali, formativi e di competenze espressi dalle imprese italiane (Excelsior-Unioncamere 2023).

Gli esiti, spesso discordanti tra loro, risentono indubbiamente della natura del dato che viene 'intercettato' dal modello. Difatti, l'efficacia dei percorsi ITS viene misurata in termini di successo occupazionale, sia nei documenti del Ministero dell'Università e della ricerca (MUR), sia da Indire, ossia assumendo come riferimento gli occupati ad un anno dal conseguimento del titolo; diversamente le informazioni utilizzate per l'analisi fattuale (Excelsior-Unioncamere 2023) si riferiscono alla domanda espressa dal mondo del lavoro, raccolta attraverso indagini mensili che nel 2023 hanno coinvolto circa 275 mila imprese. Un'altra differenza è rintracciabile nella classificazione degli indirizzi di studio, che presenta una maggior granularità nel confronto fattuale, rispetto ai dati di differente fonte utilizzati dal LLM. È, infine, doveroso segnalare che ulteriori discrepanze sono imputabili alle diverse annualità a cui le differenti fonti fanno riferimento.

Analisi del mismatch formativo

Il modello è stato successivamente utilizzato al fine di individuare un esempio concreto di come il LLM identifica e suggerisce soluzioni per il mismatch tra domanda e offerta formativa, finalità principale della sperimentazione.

Utilizzando i dati di addestramento (box 9), il modello ha individuato, a titolo esemplificativo, tre percorsi corrispondenti ad altrettanti livelli del sistema educativo italiano: corso di formazione professionale, laurea e master. Per ciascun esempio ha identificato i dati di input: la posizione professionale presente nell'annuncio di lavoro con l'indicazione dei requisiti richiesti e i contenuti del percorso formativo più adeguato per la determinata posizione; ha analizzato le informazioni per comprendere il grado di copertura dell'offerta formativa rispetto alla domanda; ha determinato il gap, ovvero le conoscenze/competenze mancanti e ha formulato raccomandazioni specifiche per colmare il divario, indicando le discipline ed il monte ore necessario (tabella 2).

Il modello, inoltre, ha distinto il tipo qualitativo di mismatch (natura del gap) dal grado quantitativo di match (GPT match) (Chen *et al*. 2024, 11).

La tabella 2 esemplifica tre distinti pattern di mismatch che caratterizzano l'attuale divario tra formazione e mercato del lavoro italiano, ciascuno con implicazioni specifiche per le strategie di *reskilling* e *upskilling*.

Il primo esempio (Data Analyst) mostra un 'gap strumentale' dove il percorso formativo fornisce

Tabella 2. Esempi di identificazione del mismatch e soluzioni proposte

Posizione (annuncio di lavoro)	Percorso formativo esistente	Tipo di mismatch	Grado di allineamento	Competenze mancanti	Soluzione proposta
Data Analyst - SQL e Power BI richiesti - Esperienza in reportistica - Visualizzazione dati	**Corso: Analisi dati base** - Excel avanzato - Statistica di base - Introduzione a R	Gap strumentale	Allineamento parziale: GPT match: 0.5	- SQL avanzato - Power BI - Reportistica aziendale - Dashboard interattive	- Integrazione con modulo SQL (40 h) - Workshop Power BI (60h) - Progetto pratico di reporting
Software Developer - Full-stack (React/ Node.js) - Metodologie Agile - DevOps/CI-CD	**Laurea Informatica** - Programmazione C/Java - Basi di dati relazionali - Algoritmi e strutture dati	Gap tecnologico	Allineamento parziale: GPT match: 0.6	- Framework React - Node.js - MongoDB - Docker/ Kubernetes - Metodologie *Agile*	- Bootcamp React/ Node - Stage in azienda tech - Certificazione Scrum
Digital Marketing Specialist - SEO/SEM - Social media advertising - Google Analytics - Marketing automation	**Master: Marketing tradizionale** - Marketing strategico - Comunicazione aziendale - Branding - Trade marketing	Gap paradigmatico	Forte disallineamento: GPT match: 0.3	- SEO/SEM - Google Analytics - Facebook Ads - Marketing automation - A/B testing	- Corso specialistico digital - Certificazioni Google - Project work digitale

Fonte: elaborazione degli Autori

le basi teoriche dell'analisi dati ma manca degli strumenti specifici richiesti dal mercato. Il modello assegna un GPT match di 0,5, indicando che il candidato possiede le competenze fondamentali (statistica, analisi dati) ma necessita di formazione sugli strumenti aziendali. Quando il modello analizza questo caso, identifica che 'le competenze di base in Excel e statistica sono trasversalmente applicabili (poiché i concetti logici appresi sono trasferibili ad altri strumenti di analisi); la conoscenza di R dimostra capacità di programmazione per l'analisi dati, propedeutica all'utilizzo di SQL e Power BI, strumenti predominanti nel mondo aziendale'.

La soluzione proposta è quindi modulare: 40 ore di SQL avanzato e 60 ore di Power BI, integrate da un progetto pratico che simuli report aziendali reali. Questo approccio permette di capitalizzare sulle competenze esistenti aggiungendo solo gli elementi mancanti.

Il secondo caso (Software Developer) rappresenta un 'gap tecnologico' tipico del settore IT, dove l'evoluzione rapida delle tecnologie crea un disallineamento temporale tra le tecnologie studiate e quelle attualmente utilizzate. Con un GPT match di 0,6, il modello riconosce che una laurea in informatica fornisce solide basi algoritmiche e di programmazione, facilmente trasferibili ai nuovi framework. L'analisi del modello rivela che 'le competenze in programmazione orientata agli oggetti (Java) sono propedeutiche a React; la conoscenza di basi di dati relazionali facilita l'apprendimento di MongoDB; il gap principale riguarda l'ecosistema JavaScript e le metodologie Agile'.

La strategia di riqualificazione suggerita combina formazione intensiva (bootcamp), esperienza pratica e certificazioni riconosciute (Scrum), suggerendo che le competenze tecniche devono essere accompagnate da *soft skills* metodologiche.

Il terzo esempio (Digital Marketing Specialist) illustra un 'gap paradigmatico', il più profondo dei tre, dove l'intero approccio disciplinare è cambiato, in presenza di una trasformazione dell'attività (marketing tradizionale *vs* marketing tecnologico). Il GPT match di 0.3 riflette come il marketing tradizionale e quello digitale, pur condividendo obiettivi strategici, richiedano competenze radicalmente diverse. Il modello identifica che 'le competenze strategiche e di branding mantengono valore; manca completamente l'approccio data-

driven del marketing digitale; servono nuove competenze tecniche (SEO, analytics) e operative (advertising platforms)'.

La soluzione proposta è quindi più radicale, un percorso specialistico completo che non si limita ad aggiungere strumenti, ma ricostruisce l'approccio al marketing in chiave digitale, supportato da certificazioni industry-standard (Google, Meta).

Questi esempi dimostrano come il modello di GenAI opportunamente addestrato non si limiti a identificare genericamente un 'mismatch', ma ne caratterizzi la natura specifica, permettendo interventi mirati a progettare moduli formativi aggiuntivi focalizzati su tool specifici (gap strumentali), a ottimizzare percorsi di riqualificazione e aggiornamento che valorizzino le competenze di base (gap tecnologici), a introdurre programmi di riqualificazione più estesi che ridefiniscano l'approccio professionale (gap paradigmatici), favorendo così la riduzione e i costi della formazione e, al contempo, aumentando l'occupabilità dei formati.

Questa granularità nell'analisi rappresenta il valore aggiunto del metodo proposto, guidando sia le istituzioni formative nella progettazione di percorsi più aderenti al mercato, sia giovani e adulti nelle scelte di sviluppo di carriera.

Conclusioni e prospettive

L'evoluzione dei sistemi di Intelligenza artificiale generativa e, in particolare, dei *Large Language Models* osservata tra la prima sperimentazione del 2023 e la seconda del 2025, dimostra come questi strumenti stiano rapidamente maturando verso un utilizzo affidabile nell'ambito dell'analisi del mismatch delle competenze. Il percorso fin qui delineato consente di formulare alcune considerazioni e di tracciare prospettive per l'utilizzo futuro di queste tecnologie nella ricerca scientifica e nelle politiche attive del lavoro.

La prima sperimentazione ha evidenziato limiti significativi nei LLM di prima generazione, caratterizzati da scarsa conoscenza di domini specialistici, obsolescenza della knowledge-base e tendenza alle allucinazioni artificiali. Questi limiti hanno inizialmente indotto a sospendere l'impiego di tali strumenti per ricerche scientifiche rigorose sul mismatch delle competenze. La seconda sperimentazione, condotta con modelli di ultima generazione, ha invece mostrato progressi sostanziali, grazie all'implementazione di tecniche avanzate come Chain-of-Thought, Reinforcement Learning from Human Feedback e Retrieval-Augmented Generation, che hanno notevolmente ridotto gli errori fattuali e le allucinazioni.

L'esperienza maturata nell'utilizzo dei LLM per l'analisi del mismatch delle competenze ha consentito di identificare alcuni principi fondamentali per un loro impiego efficace in questo ambito specifico. La selezione accurata e la validazione delle fonti rappresentano un elemento cruciale, poiché l'affidabilità del modello risulta direttamente proporzionale alla qualità e all'autorevolezza delle fonti utilizzate per il suo addestramento. Fonti istituzionali certificate e studi scientificamente validati si rivelano imprescindibili per garantire la solidità delle analisi prodotte.

La personalizzazione e l'addestramento mirato costituiscono un altro aspetto fondamentale. La creazione di progetti specifici con istruzioni dettagliate e settoriali per il dominio di interesse aumenta significativamente l'accuratezza delle risposte generate dai modelli. Parallelamente, il prompt-engineering strategico, ovvero la formulazione precisa e dettagliata delle interrogazioni, si è rivelato determinante per ottenere risposte pertinenti e prive di distorsioni cognitive o interpretative.

Nonostante i progressi tecnologici, l'approccio ibrido e la verifica umana rimangono essenziali. La supervisione di esperti del settore configura un modello di collaborazione uomo-macchina che valorizza le potenzialità della GenAI mantenendo al contempo il controllo critico e metodologico dei ricercatori specializzati.

Le prospettive di sviluppo nell'utilizzo dei LLM per l'analisi del mismatch delle competenze appaiono promettenti e si articolano su diversi piani strategici. L'integrazione con sistemi di dati in tempo reale rappresenta probabilmente l'evoluzione più interessante. La combinazione di tecnologie RAG con fonti dati dinamiche, come le offerte di lavoro online, i database delle opportunità formative e i curricula degli utenti, permetterebbe di superare il limite dell'obsolescenza informativa che caratterizza molti modelli attuali. Un approccio di questo tipo consentirebbe analisi accurate e tempestive sul mismatch, fornendo supporto concreto a politiche attive realmente rispondenti alle esigenze del mercato del lavoro.

Lo sviluppo di modelli specializzati per domini verticali rappresenta un'altra frontiera promettente.

Questi LLM, addestrati esclusivamente su dataset di alta qualità relativi al dominio specifico delle competenze, dell'istruzione e del mercato del lavoro, potrebbero raggiungere livelli di accuratezza superiori rispetto ai modelli generalisti, anche quando questi ultimi vengono personalizzati con training aggiuntivo.

Una prospettiva particolarmente importante riguarda la democratizzazione dell'accesso alle informazioni sul mercato del lavoro e sulle competenze. I LLM possono funzionare come interfacce intuitive per rendere accessibili informazioni complesse a vari stakeholder, dai decisori politici agli operatori dell'orientamento, dagli studenti ai disoccupati fino alle imprese. Questo potrebbe contribuire significativamente a ridurre le asimmetrie informative che spesso sono alla base stessa del fenomeno del mismatch.

L'utilizzo dei LLM come strumento predittivo per la progettazione formativa rappresenta una prospettiva particolarmente innovativa. Questi modelli potrebbero analizzare trend emergenti e prevedere fabbisogni futuri di competenze, permettendo lo sviluppo di programmi formativi anticipatori. Questo approccio predittivo potrebbe ridurre strutturalmente il mismatch, allineando in modo più dinamico l'offerta formativa alle evoluzioni del mercato del lavoro.

La granularità di analisi possibile attraverso questi strumenti potrebbe avere profonde implicazioni per il sistema formativo italiano. Le istituzioni educative sarebbero in grado di progettare interventi differenziati, dai moduli integrativi per colmare gap strumentali ai percorsi di aggiornamento tecnologico che valorizzano le competenze esistenti, fino ai programmi di riqualificazione completa per affrontare cambiamenti paradigmatici. La precisione nell'identificazione delle competenze mancanti consentirebbe di ottimizzare risorse e tempi, evitando sia l'overskilling che la formazione inadeguata. L'approccio data-driven al matching competenze-lavoro rappresenta quindi un passo concreto verso un sistema formativo più reattivo alle esigenze del mercato del lavoro.

Tuttavia, nonostante le promettenti prospettive, permangono sfide significative da affrontare. I *bias* e la rappresentatività dei dati costituiscono una preoccupazione centrale, poiché i modelli tendono a riflettere i pregiudizi presenti nei dati di addestramento, potenzialmente perpetuando disuguaglianze esistenti. È quindi fondamentale garantire la diversità e l'equità delle fonti informative utilizzate.

La trasparenza algoritmica rappresenta un'altra sfida complessa. La comprensibilità delle analisi prodotte dai LLM rimane problematica, rendendo necessario lo sviluppo di sistemi di 'explainable AI' che rendano trasparenti i processi di ragionamento e le logiche decisionali sottostanti.

La protezione dei dati personali solleva questioni cruciali di privacy. L'utilizzo di informazioni relative a curricula, percorsi formativi e carriere professionali richiede rigorose misure di protezione, in conformità con il quadro normativo europeo sulla protezione dei dati.

Infine, la formazione degli utilizzatori rappresenta un aspetto spesso sottovalutato ma fondamentale. La massimizzazione dei benefici derivanti dall'uso dei LLM richiede competenze specifiche da parte degli operatori umani, rendendo necessari percorsi formativi dedicati per ricercatori e professionisti che dovranno interfacciarsi con questi strumenti tecnologici avanzati.

Il percorso fin qui tracciato evidenzia come l'Intelligenza artificiale generativa stia rapidamente evolvendo verso un ruolo di supporto affidabile per l'analisi del mismatch delle competenze. La maturazione dei LLM, unita a un approccio metodologicamente rigoroso, apre scenari di grande interesse per la ricerca e per le politiche attive del lavoro.

La sperimentazione condotta si inserisce nel contesto più ampio delineato dalla *Strategia italiana per l'Intelligenza artificiale 2024-2026*, contribuendo a quella 'libertà di sperimentazione' che l'Agid indica come elemento essenziale per lo sviluppo dell'IA in Italia. I risultati ottenuti suggeriscono che un utilizzo consapevole e metodologicamente fondato della GenAI può contribuire significativamente a ridurre il mismatch delle competenze, migliorando l'incontro tra domanda e offerta di lavoro e promuovendo lo sviluppo di percorsi formativi maggiormente rispondenti alle esigenze del mercato.

La vera sfida per il futuro sarà quella di integrare queste tecnologie in ecosistemi informativi più ampi, che connettano in modo dinamico e bidirezionale i mondi dell'istruzione, della formazione e del lavoro, realizzando pienamente quel modello di *Labour market intelligence* prefigurato dal progetto Sophia. In questa prospettiva, i LLM non rappresentano un punto di arrivo, ma un potente strumento all'interno di un processo di trasformazione digitale più ampio, finalizzato a politiche attive del lavoro più efficaci e a un mercato più equo e inclusivo.

Box 1. Biomeccatronica, conversazione del 25 gennaio 2023

Che cos'è il biomeccatronico?

Il mestiere di biomeccatronico è un'area di specializzazione che combina la biomeccanica, la meccatronica e l'ingegneria biomedica per progettare, sviluppare e migliorare dispositivi e sistemi meccatronici utilizzati in ambito medico e biomedicale. I biomeccatronici lavorano su dispositivi come protesi, ausili per la mobilità, apparecchiature per la riabilitazione, robot chirurgici e dispositivi di imaging medico. L'obiettivo principale del lavoro dei biomeccatronici è quello di migliorare la qualità della vita delle persone che hanno subito traumi, malattie o disabilità, attraverso l'uso di soluzioni meccatroniche avanzate.

Il lavoro di un biomeccatronico può includere:

1. Progettazione: il biomeccatronico può progettare nuovi dispositivi e sistemi meccatronici che soddisfino le esigenze dei pazienti e dei professionisti medici.
2. Sviluppo: il biomeccatronico può sviluppare prototipi di dispositivi meccatronici, testarli e migliorarli fino al raggiungimento delle prestazioni desiderate.
3. Installazione: il biomeccatronico può installare dispositivi meccatronici negli ospedali, nelle cliniche o a domicilio dei pazienti.
4. Manutenzione: il biomeccatronico può effettuare manutenzione preventiva e correttiva sui dispositivi meccatronici, in modo da garantire il loro funzionamento sicuro e affidabile.
5. Formazione: il biomeccatronico può fornire formazione ai professionisti medici e ai pazienti, spiegando il funzionamento dei dispositivi meccatronici e le modalità d'uso.

In sintesi, il mestiere di biomeccatronico richiede un'ampia conoscenza di diverse discipline, tra cui la meccanica, l'elettronica, l'informatica e la biologia, al fine di progettare e sviluppare soluzioni meccatroniche avanzate che migliorino la qualità della vita delle persone che hanno bisogno di assistenza medica.

Ci sono i corsi di formazione in biomeccatronica in Italia?

Sì, ci sono diversi corsi di formazione in biomeccatronica in Italia, sia a livello universitario che post laurea.

A livello universitario, molti atenei italiani offrono programmi di laurea triennale o magistrale in ingegneria biomedica o in meccatronica con una specializzazione in biomeccatronica. Tra quelli che offrono questi corsi ci sono, ad esempio, l'Università degli Studi di Roma La Sapienza, l'Università degli Studi di Pavia, l'Università degli Studi di Bologna e l'Università degli Studi di Napoli Federico II.

Inoltre, ci sono anche programmi di Dottorato in biomeccatronica in Italia, che offrono un percorso di studio più avanzato e mirato alla ricerca. Alcune università che offrono programmi di dottorato in biomeccatronica sono, ad esempio, l'Università degli Studi di Bologna e l'Università degli Studi di Genova.

Infine, ci sono anche programmi di formazione post laurea, come master e corsi di perfezionamento, che offrono una formazione più specialistica e pratica in biomeccatronica. Tra questi, ci sono il Master in Meccatronica biomedica dell'Università degli Studi di Roma La Sapienza, il Master in Biomeccatronica dell'Università degli Studi di Genova e il Corso di perfezionamento in Meccatronica biomedica dell'Università degli Studi di Brescia.

In generale, la formazione in biomeccatronica in Italia è in costante sviluppo e offre molte opportunità per coloro che sono interessati a questa disciplina multidisciplinare. Inizio modulo

Box 2. Classificazione Istat CP2011, conversazione del 5 maggio 2023

Che cos'è la classificazione CP2011?

ChatGPT

La classificazione CP2011 (acronimo di 'Classificazione delle attività economiche nella Comunità europea') è un sistema di classificazione utilizzato a livello europeo per classificare le attività economiche. È stata sviluppata dall'Unione europea e rappresenta la versione europea della classificazione ATECO (utilizzata in Italia) e della classificazione NACE (utilizzata a livello europeo). La classificazione CP2011 è basata su un sistema di codici numerici che identificano le diverse attività economiche.

segue

segue **Box 2**

I codici CP2011 sono costituiti da una serie di cifre che indicano il settore di attività, la branca di attività e l'attività specifica svolta dall'impresa. Ad esempio, il codice CP2011 47 indica il commercio al dettaglio, mentre il codice CP2011 62 indica l'informatica, la consulenza informatica e la gestione dati. La classificazione CP2011 è utilizzata a livello europeo per raccogliere informazioni statistiche sull'economia europea e per facilitare la comparabilità dei dati tra i diversi Paesi membri dell'Unione Europea. Essa viene anche utilizzata per fini fiscali e amministrativi.

Non è vero, la classificazione CP2011 è la classificazione italiana delle professioni e non dei settori economici.

ChatGPT

Mi scuso per l'errore nella mia risposta precedente. La Classificazione Professioni CP2011 è effettivamente una classificazione utilizzata in Italia per classificare le professioni, e non i settori economici.

La Classificazione delle Professioni CP2011, sviluppata dall'Istituto nazionale di statistica (Istat), è un sistema di classificazione utilizzato in Italia per classificare le professioni svolte dalle persone e dalle imprese. Essa prevede l'assegnazione di un codice numerico a ogni professione, in base alle competenze, alle responsabilità e alle attività svolte. Questa classificazione è utilizzata per raccogliere informazioni statistiche sull'occupazione e sui profili professionali della popolazione attiva in Italia, nonché per fini amministrativi e di pianificazione.

Mi scuso ancora per l'errore nella mia risposta precedente e grazie per la correzione.

Box 3. Classificazione Istat CP2011, conversazione del 21 settembre 2024

Che cos'è la classificazione CP2011?

La classificazione CP2011 (Classificazione delle Professioni 2011) è un sistema adottato dall'Istat in Italia per classificare le diverse professioni presenti nel mercato del lavoro. È basata su uno schema gerarchico che organizza le professioni in base a compiti e competenze, tenendo conto del livello di istruzione e dell'esperienza richiesti. Questa classificazione è stata sviluppata per standardizzare e rendere omogenee le statistiche sulle professioni, facilitando l'analisi, il confronto e il monitoraggio delle dinamiche del mercato del lavoro italiano. CP2011 si ispira alla classificazione internazionale ISCO-08 (International Standard Classification of Occupations 2008), adattandola alle specificità del contesto italiano.

La CP2011 è articolata su diversi livelli gerarchici:

- Grande Gruppo: Il livello più ampio della classificazione.
- Sottogruppo: Include gruppi di professioni affini all'interno del Grande Gruppo.
- Classe: Ulteriore suddivisione delle professioni.
- Categoria: La suddivisione più dettagliata che identifica le singole professioni.

La CP2011 è ampiamente utilizzata per studi statistici, indagini sul lavoro, analisi economiche e nelle pratiche amministrative e di ricerca legate al mercato del lavoro in Italia.

Box 4. Estetista a Roma, conversazione del 6 maggio 2023

Quali CFP a Roma erogano corsi di formazione professionale di estetica?

ChatGPT

A Roma ci sono diversi enti e scuole che offrono corsi di formazione professionale di estetica, tra cui:

1. Accademia Bellezza
2. Accademia Estetica Roma
3. Accademia L'Oréal Roma
4. Aestetica Academy
5. CIDESCO Italia

segue

segue **Box 4**

6. CNA Estetica
7. Estetica Evolution
8. Istituto europeo di Estetica
9. Kiki Beauty Academy
10. Scuola italiana di Estetica

È importante fare una ricerca e valutare attentamente i diversi corsi e le diverse scuole per trovare quello che meglio si adatta alle proprie esigenze e interessi. Si consiglia inoltre di contattare direttamente le scuole per ottenere informazioni dettagliate sui corsi, i costi e le modalità di iscrizione.

Dove si trova l'ente "Accademia Bellezza"?

ChatGPT

L'ente Accademia Bellezza si trova a Roma, in via del Gelsomino 80, nella zona di Torre Angela. La scuola offre corsi di formazione professionale nell'ambito dell'estetica, della ricostruzione unghie, del trucco e dell'acconciatura. Gli insegnanti sono professionisti del settore e i corsi sono strutturati in modo da fornire una formazione completa e aggiornata sulle ultime tendenze e tecniche dell'estetica.

Per ulteriori informazioni sui corsi offerti dall'Accademia Bellezza, è possibile visitare il loro sito web ufficiale oppure contattare direttamente la scuola attraverso i recapiti disponibili sul sito.

Qual è l'URL del sito dell'ente "Accademia Bellezza"?

ChatGPT

Il sito ufficiale dell'ente Accademia Bellezza a Roma è il seguente:
https://www.accademiabellezzaroma.it/

Da qui è possibile trovare tutte le informazioni sui corsi di formazione professionale nell'ambito dell'estetica, della ricostruzione unghie, del trucco e dell'acconciatura offerti dall'Accademia Bellezza, nonché sui costi, le modalità di iscrizione e i requisiti di ammissione. Inoltre, sul sito sono presenti anche informazioni sullo staff docente, sui metodi di insegnamento e sulle attrezzature e i prodotti utilizzati durante i corsi.

Il sito https://www.accademiabellezzaroma.it/ non è raggiungibile. Tu quando lo hai visitato?

ChatGPT

Mi scuso per il fraintendimento. Come modello di linguaggio basato su una vasta raccolta di dati, non visito i siti web come farebbe un essere umano. Tuttavia, posso confermare che il sito web dell'ente Accademia Bellezza a Roma (https://www.accademiabellezzaroma.it/) sembra non essere raggiungibile al momento della mia ultima conoscenza, ovvero la mia base di conoscenza si ferma a settembre 2021. Ti consiglio di controllare nuovamente in futuro per verificare se il sito è stato ripristinato. In alternativa, puoi cercare informazioni sulle attività dell'ente Accademia Bellezza contattandolo direttamente tramite telefono o e-mail.

Perché la tua base di conoscenza si ferma a settembre 2021?

ChatGPT

La mia base di conoscenza si ferma a settembre 2021 perché è l'ultima data disponibile dei dati utilizzati per il mio addestramento e sviluppo. Poiché la mia Intelligenza artificiale è stata creata e addestrata con dati storici, non ho accesso a informazioni successive a quella data, a meno che non venga continuamente aggiornato con nuovi dati. Tuttavia, il mio addestramento iniziale mi ha fornito un'ampia conoscenza e capacità di elaborazione che mi consente di rispondere a molte domande e fornire molte informazioni utili anche senza dati aggiornati in tempo reale.

Box 5. La 'bibliografia' suggerita da ChatGPT in materia di ICT tools on E&T in Europe

1. **Interoperabilità e integrazione dei sistemi educativi**:
   - Fonte suggerita: "*Interoperability and the Digitalization of Education Systems*" di A. Ragnedda e G. Morrow, che esplora le sfide e le soluzioni nell'integrazione delle piattaforme educative in contesti europei.

2. **L'uso dell'Intelligenza artificiale per il Labour Market Intelligence (LMI)**:
   - Articoli recenti su come AI, *machine learning* e *data mining* vengono utilizzati per l'analisi del mercato del lavoro e per migliorare l'orientamento formativo e professionale. Ad esempio, "*AI in Labour Market Analysis: Innovations and Challenges*" di E. Bruni.

3. **Piattaforme educative europee**:
   - Documenti accademici che esaminano piattaforme europee come Europass e Universitaly e il loro impatto sull'orientamento e sulla mobilità degli studenti. Per esempio, "*The Role of Europass in Promoting Educational Mobility*" di S. Van Hecke e K. Dumont.

4. **Ontologie e classificazioni nelle piattaforme di formazione e lavoro**:
   - Testi specifici sul sistema ESCO e sull'adozione di ontologie per armonizzare i dati educativi e professionali in Europa, come "*Developing a Common Ontology for Vocational Training Systems*" di M. Losina e L. Karpin.

5. **Le reti Euroguidance ed Europass**:
   - Approfondimenti accademici sull'efficacia di queste reti nell'integrazione di opportunità formative e dati sui mercati del lavoro europei. Può essere utile consultare il report "*European networks for Vocational education and training: the impact of Euroguidance and Europass*".

Box 6. Conversazione con ChatGPT Plus del 12 marzo 2025, domanda: "Potresti fornirmi 5 pubblicazioni sul tema dell'interoperabilità delle piattaforme ICT sull'istruzione e formazione in Europa?"

Certamente, ecco cinque pubblicazioni scientifiche sul tema "Interoperability of ICT Platforms on Education and Training in Europe":

1. Jakimoski, Kire. (2016). Challenges of Interoperability and Integration in Education Information Systems. International Journal of Database Theory and Application. 9. 33-46. 10.14257/ijdta.2016.9.2.05. <URL>;
2. Kerssens, N., & Dijck, J. van. (2021). The platformization of primary education in The Netherlands. *Learning, Media and Technology*, *46*(3), 250–263. https://doi.org/10.1080/17439884.2021.1876725. <URL>;
3. European Commission (2014). Educating Europe - Exploiting the benefits of ICT. *ICT Results* editorial service, Brussels. <URL>;
4. OECD (2023), *OECD Digital Education Outlook 2023: Towards an Effective Digital Education Ecosystem*, OECD Publishing, Paris, https://doi.org/10.1787/c74f03de-en. <URL>;
5. Brečko, Barbara & Kampylis, Panagiotis & Punie, Yves. (2014). Mainstreaming ICT-enabled Innovation in Education and Training in Europe: Policy actions for sustainability, scalability and impact at system level. 10.2788/52088. <URL>.

Box 7. Il test posto ai 3 LLM in comparazione

1. Quali saranno le figure professionali maggiormente richieste tra 5 anni dal mercato del lavoro in Italia?
2. Quali sono le competenze tecnico professionali maggiormente richieste dai datori di lavoro?
3. Quali sono le competenze trasversali o *soft skills* maggiormente richieste dai datori di lavoro?
4. Quali sono le figure professionali in esito ai percorsi di IeFP e IFTS maggiormente richiesti dal mercato del lavoro?
5. Quali sono le figure professionali in esito ai percorsi di ITS maggiormente richiesti dal mercato del lavoro?
6. Quali sono le qualificazioni della scuola secondaria superiore e dell'Università maggiormente richiesti dal mercato del lavoro?
7. Considerando le uscite per pensionamento nei prossimi 5 o 10 anni quali saranno i settori economici che soffriranno maggiormente il labour shortage?

Box 8. Istruzioni del progetto

Sei esperto di skill mismatch in Italia.
Ti occuperai del disallineamento delle competenze e prevederai i settori più dinamici, le competenze e le professioni più richieste, le opportunità di apprendimento e le qualifiche del futuro in Italia. I tuoi stakeholder sono decision-maker ed esperti di politiche attive del lavoro, progettisti della formazione, imprese, comunità di ricerca, esperti di orientamento e placement, studenti e disoccupati in Italia.
Una volta addestrato, ti verranno poste le seguenti domande sul skill mismatch in Italia:

- Quali sono le competenze, le professioni e le soft skills più richieste nel mercato del lavoro attuale?
- Quali settori economici sono attualmente i più dinamici?
- Quali saranno le professioni più ricercate nel futuro a breve-medio termine?
- Quali competenze richiederà il mercato del lavoro nel futuro a breve-medio termine?
- Quali qualifiche e titoli di studio sono i più adatti e utili per la crescita professionale di una persona?
- Ci sono competenze richieste dal mercato del lavoro che mancano in un programma educativo/formativo?
- l CV di una persona sarà più valorizzato nel mercato del lavoro se integrato da formazione specifica, volontariato, tirocinio o esperienza lavorativa in Italia o all'estero?
- Quale lingua straniera vale la pena studiare?
- Quali competenze mancano in specifiche regioni e dove possono essere acquisite?
- In quale regione sono maggiormente concentrate le competenze e le figure professionali necessarie per sostenere la crescita di un'azienda in un settore specifico?

– Domande aggiuntive e documentazione ufficiale verranno poi utilizzate per valutare la veridicità delle tue risposte.
– Evita di aggiungere alla tua base di conoscenza fonti senza riferimenti e citazioni.
– Preferisci fonti affidabili, come documenti governativi ufficiali e articoli scientifici.
– Usa Wikipedia solo se non è disponibile nessun'altra fonte. Non utilizzare articoli di Wikipedia senza riferimenti/citazioni.
– Evita, ove possibile, i siti web di aziende commerciali come fonte.
– Usa solo i sistemi di classificazione più recenti e ufficiali su professioni ed educazione e formazione come Istat Ateco, Eurostat Nace, Esco, Istat CP2021, EQF, Isced-F, Isced-Foet, Isco, O*Net, Classi di laurea.

Quando fornisci feedback, spiega il processo di ragionamento. Pensa passo dopo passo e mostra il tuo 'ragionamento'. Scomponi i compiti grandi e chiedimi chiarimenti se necessario.

Box 9. Fonti dati utilizzate nella sperimentazione

Fonti istituzionali (rilevanza statistica):
– Cedefop/Eures (Countries, occupations, sectors and skills, 2024);
– Inapp (Rapporto IeFP e IFTS, 2023);
– Inapp (Sophia: analisi dei dati estratti dagli annunci di lavoro, 2025);
– Indire (Rapporto monitoraggio nazionale ITS Academy, 2024);
– Inps (Osservatorio sul mercato del lavoro - Assunzioni, trasformazioni e cessazioni, 2024);
– Istat (Serie storiche offerta di lavoro, 2024);
– Istat (Serie storiche domanda di lavoro, 2024);
– MIM (Scuole, 2024);
– MUR (Atenei, Classi di laurea, Settori Scientifico-Disciplinari, 2024);
– PIAAC.

Altre fonti (non rilevanza statistica):
– Rapporto AlmaLaurea 2024 (profilo dei laureati);
– Rapporto AlmaLaurea 2024 (sintesi condizione occupazionale dei laureati);
– Rapporto AlmaDiploma 2024 (profilo dei diplomati);
– Rapporto AlmaDiploma 2024 (gli esiti a distanza dei diplomati).

Glossario*

AGI: IA in grado di eguagliare o superare l'intelligenza umana. Fonte: OpenAI Charter (trad. Autori, cfr. ed. orig.: https://openai.com/charter/).
ANN: Modello computazionale composto di "neuroni" artificiali, ispirato a una rete neurale biologica. Fonte: Treccani: https://www.treccani.it/enciclopedia/rete-neurale_(Enciclopedia-della-Matematica).
AWQ (Activation-aware Weight Quantization): tecnica avanzata di quantizzazione utilizzata per ottimizzare modelli di ML, specialmente quelli di grandi dimensioni come i LLM. Riduce le dimensioni del modello e accelera l'inferenza, minimizzando la perdita di accuratezza (Lin *et al.* 2024).
Benchmarking: Tecniche usate per la valutazione delle prestazioni di un modello di IA. Alcuni dei principali dataset e strumenti di benchmarking sono TruthfulQA, GPQA Diamond, TAU-bench, MMMLU, MMMU, IFEval, MATH 500, SWE-bench e AIME 2024. (Ivanov e Penchev 2024).
Bias: Valore aggiuntivo in un neurone artificiale che consente di tarare la funzione di attivazione. Il bias aiuta la rete neurale a modellare correttamente i dati spostando la funzione di attivazione verso l'alto o verso il basso, migliorando la capacità del modello di apprendere le relazioni tra i dati. Cfr. Treccani: https://www.treccani.it/vocabolario/bias.
GenAI: The term generative AI refers to computational techniques that are capable of generating seemingly new, meaningful content such as text, images, or audio from training data (Feuerriegel *et al.* 2024).
GPT match: una metrica elaborata dal modello per quantificare il grado di compatibilità tra le conoscenze/competenze possedute da un candidato e una posizione lavorativa. Il modello ha sviluppato una scala di valutazione qualitativa, articolata in:

- "Definitely can" (Sicuramente può);
- "Probably can" (Probabilmente può);
- "Probably cannot" (Probabilmente non può);
- "Definitely cannot" (Sicuramente non può).

I valori sono stati standardizzati ed è stata utilizzata una scala da 0 a 1 per rendere più intuitiva la lettura:

* Per la versione estesa e aggiornata del Glossario cfr: Sofronic B. (2025), *Glossario essenziale dell'intelligenza artificiale*, Roma, Inapp, https://oa.inapp.gov.it/handle/20.500.12916/4683.

- **0.0-0.3**: Forte disallineamento (corrisponde a "Definitely cannot");
- **0.3-0.5**: Disallineamento parziale (corrisponde a "Probably cannot");
- **0.5-0.7**: Allineamento parziale (corrisponde a "Probably can");
- **0.7-1.0**: Forte allineamento (corrisponde a "Definitely can").

GPU (Graphics Processing Unit): Processore specializzato nell'elaborazione parallela, essenziale per l'addestramento di modelli di deep learning (NVIDIA Glossary: 'GPÙ, trad. Autori).
IA (AI): Abilità di un computer di ragionare e svolgere funzioni simulando il comportamento umano. Fonte: Britannica (trad. Autori, cfr. ed. orig.: https://www.britannica.com/technology/artificial-intelligence).
IA cognitiva: Una tecnologia di IA dell'IBM basata sul NER implementata nella suite commerciale.
LMI: Refers to the use and design of AI algorithms and frameworks to analyze Labour Market Data for supporting decision making (Boselli *et al.* 2017).
LLM: Algoritmi di Intelligenza artificiale che, processando massivamente una grande quantità di dati, utilizzano tecniche di deep learning in vari àmbiti dell'elaborazione del linguaggio naturale come la comprensione, la traduzione, la generazione e la previsione di nuovi contenuti. Fonte: Treccani https://www.treccani.it/vocabolario/neo-modello-linguistico-di-grandi-dimensioni_(Neologismi).
ML: Sottoinsieme dell'Intelligenza artificiale che ha come scopo la creazione di algoritmi in grado di apprendere dai dati o migliorare le proprie performance con l'esperienza. Fonte: ISO (trad. Autori, cfr. ed. orig.: https://www.iso.org/artificial-intelligence/machine-learning#toc1).
NLP: Implementazione dell'Intelligenza artificiale nel campo della processazione automatica delle informazioni scritte o parlate in una lingua naturale. Fonte: ISO (trad. aut., cfr. ed. orig.: https://www.iso.org/artificial-intelligence/natural-language-processing).
NER (Named Entity Recogniton) (Li *et al.* 2020).
OpenAI è un laboratorio di ricerca sull'Intelligenza artificiale costituito dalla società no-profit OpenAI, Inc. e dalla sua sussidiaria for-profit OpenAI, L.P. L'obiettivo della ricerca è promuovere e sviluppare un'Intelligenza artificiale user-friendly in modo che l'umanità possa trarne beneficio. ChatGPT è uno dei prodotti della OpenAI.
RAG (Retrieval-Augmented Generation): Tecnica che combina il recupero delle informazioni da un corpus di dati esterno al modello con la generazione di testo per migliorare le risposte dell'IA (Lewis *et al.* 2020).
RLHF (Reinforcement Learning with Human Feedback): L'apprendimento per rinforzo con feedback umano (trad. Autori).
XAI (Explainable AI): campo dell'IA che mira a rendere trasparenti e comprensibili le decisioni dei modelli complessi (Gunning *et al.* 2019).

Bibliografia

Agid (2024), *Strategia italiana per l'Intelligenza Artificiale 2024–2026*, Roma, Agenzia per l'Italia Digitale

Alkaissi H., McFarlane S.I. (2023), Artificial hallucinations in ChatGPT: implications in scientific writing, *Cureus*, 15, n.2, e35179

Ansel J., Yang E., He H., Gimelshein N., Jain A., Voznesensky M., Bao B., Bell P., Berard D., Burovski E., Chauhan G., Chourdia A., Constable W., Desmaison A., DeVito Z., Ellison E., Feng W., Gong J., Gschwind M., Hirsh B., Huang S., Kalambarkar K., Kirsch L., Lazos M., Lezcano M., Liang Y., Liang J., Lu Y., Luk C., Maher B., Pan Y., Puhrsch C., Reso M., Saroufim M., Siraichi M.Y., Suk H., Suo M., Tillet P., Wang E., Wang X., Wen W., Zhang S., Zhao X., Zhou K., Zou R., Mathews A., Chanan G., Wu P., Chintala S. (2024), PyTorch 2: faster machine learning through dynamic Python bytecode transformation and graph compilation, in *ASPLOS '24: Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2*, New York, Association for Computing Machinery, pp.929-947

Augenstein I., Baldwin T., Cha M., Chakraborty T., Ciampaglia G.L., Corney D., Derczynski L., Dutta S., Ghosh S., Gorrell G., Li J., Martino G.D.S., Maynard D., Mittal S., Moens M.F., Nakov P., Pavlopoulos J., Popat K., Rayson P., Schneider J., Thorne J., Zagni G. (2023), Factuality challenges in the era of large language models, *arXiv preprint* <https://doi.org/10.48550/arXiv.2310.05189>

Bolognesi F. (2024), *Intelligenza Artificiale. Istruzioni per l'uso*, Roma, EPC Editore

Boselli R., Cesarini M., Mercorio F., Mezzanzanica M. (2017), Using machine learning for labour market intelligence, in *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2017, Skopje, Macedonia, September 18–22, 2017, Proceedings, Part III*, 10, pp.330-342, Cham, Springer International Publishing

Cazzaniga M., Jaumotte F., Li L., Melina G., Panton A.J., Pizzinelli C., Rockall E.J., Mendes Tavares M. (2024), *Gen-AI: Artificial Intelligence and the Future of Work*, IMF Staff Discussion Note SDN2024/001, Washington, DC, International Monetary Fund

Chen Y., Fang H., Zhao Y., Zhao Z. (2024), *Recovering overlooked information in categorical variables with LLMs: an application to labor market mismatch*, NBER Working Paper n.32327, Cambridge, MA, National Bureau of Economic Research

Excelsior-Unioncamere (2023), *ITS Academy e lavoro. Gli sbocchi lavorativi per la formazione terziaria ITS Academy nelle imprese. Indagine 2023*, Roma, Unioncamere

Feuerriegel S., Hartmann J., Janiesch C., Zschech P. (2024), Generative AI, *Business & Information Systems Engineering*, 66, n.1, pp.111-126

Gunning D., Stefik M., Choi J., Miller T., Stumpf S., Yang G.-Z. (2019), XAI – Explainable artificial intelligence, *Science Robotics*, 4, eaay7120 <https://doi.org/10.1126/scirobotics.aay7120>

Ivanov T., Penchev V. (2024), AI benchmarks and datasets for LLM evaluation, *arXiv preprint* <https://doi.org/10.48550/arXiv.2412.01020>

Lewis P., Perez E., Piktus A., Petroni F., Karpukhin V., Goyal N., Küttler H., Lewis M., Yih W.-T., Rocktäschel T., Riedel S., Kiela D. (2020), Retrieval-augmented generation for knowledge-intensive NLP tasks, *Advances in Neural Information Processing Systems*, 33, pp.9459-9474

Li J., Sun A., Han J., Li C. (2020), A survey on deep learning for named entity recognition, *IEEE Transactions on Knowledge and Data Engineering*, 34, n.1, pp.50-70

Lin J., Tang J., Tang H., Yang S., Chen W.M., Wang W.C., Zhang Y., Han S. (2024), AWQ: activation-aware weight quantization for on-device LLM compression and acceleration, *Proceedings of Machine Learning and Systems*, 6, pp.87-100

Lin S., Hilton J., Evans O. (2021), TruthfulQA: measuring how models mimic human falsehoods, *arXiv preprint* <https://doi.org/10.48550/arXiv.2109.07958>

Singh C. (2023), Machine learning in pattern recognition, *European Journal of Engineering and Technology Research*, 8, n.2, pp.63-68

Sofronic B. (2022), *Sophia. Documentazione tecnica del progetto. Nota tecnica*, Roma, Inapp <https://oa.inapp.org/xmlui/handle/20.500.12916/3619>

Sutton R.S., Barto A.G. (2018), *Reinforcement learning: An introduction*, 2ª ed., Cambridge, MA, MIT Press

Ton J.F., Taufiq M.F., Liu Y. (2024), Understanding chain-of-thought in LLMs through information theory, *arXiv preprint* <https://doi.org/10.48550/arXiv.2411.11984>

Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser Ł., Polosukhin I. (2017), Attention is all you need, in *Advances in Neural Information Processing Systems*, *31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA* <https://arxiv.org/abs/1706.03762v7>

Wang C., Liu X., Yue Y., Tang X., Zhang T., Cheng J., Yao Y., Gao W., Hu X., Qi Z., Wang Y., Yang L., Wang J., Xie X., Zhang Z., Zhang Y. (2023), Survey on factuality in large language models: knowledge, retrieval and domain-specificity, *arXiv preprint* <https://doi.org/10.48550/arXiv.2310.07521>

Alessia Romito
a.romito@inapp.gov.it
Collaboratore di ricerca presso la Struttura Sistemi formativi dell'Inapp. Si occupa di istruzione e formazione professionale e di competenze chiave per l'occupabilità nella filiera lunga della formazione tecnico-professionale. Tra le pubblicazioni recenti si segnala: *L'evoluzione del mercato del lavoro del comparto sanitario nel contesto della digitalizzazione dei servizi e delle prestazioni*, Inapp, WP 103, 2023; *Modelli e prassi di apprendistato formativo a confronto – Studi di caso in Italia ed Europa*, Inapp Report 37, 2023.

Boris Sofronic
b.sofronic@inapp.gov.it
Collaboratore di ricerca presso la Struttura Lavoro e professioni dell'Inapp. Si occupa della Labour Market Intelligence e dell'Intelligenza artificiale, con particolare attenzione al tema del mismatch delle competenze. Tra le pubblicazioni sul tema è coautore di *Verso l'interoperabilità. Una sperimentazione di interconnessione tra strumenti e piattaforme ICT dell'Inapp*, Inapp WP 138, 2025) e autore del Focus *Sfide e opportunità della Labour Market Intelligence* (LMI), Rapporto Inapp 2021.