

UNIVERSITÀ DEGLI STUDI DI SALERNO

DIPARTIMENTO DI INFORMATICA

Progetto di Fondamenti di Intelligenza Artificiale



La Bugia degli LLM - Perché le macchine non possono pensare

*Lo sviluppatore sta cambiando, non scomparendo.
Lo sviluppatore è immortale.*

Studente

Antonio Caiazzo

Matricola: 0512117751

ANNO ACCADEMICO 2024/2025

Indice

1	Introduzione	1
1.1	Come nasce questa relazione	1
1.2	Argomenti della relazione	3
2	Cos'è il pensiero? L'umano vs la macchina	4
2.1	La natura complessa del pensiero umano	6
2.2	Perché le macchine non possono pensare	6
3	La filosofia del pensiero e della creazione	7
3.1	Dove gli LLM falliscono nel creare qualcosa di nuovo	10
4	Principali rischi legati agli LLM	11
4.1	Rischi nell'uso e nell'implementazione degli LLM	11
5	L'etica nell'AI	15
5.1	Il trolley problem applicato all'AI	16
5.2	Bias e complessità etica	16
5.3	Un'etica è programmabile?	17
6	Perché l'AI non sostituisce lo sviluppatore	18
6.1	Gli LLM simulano ma non comprendono	18
6.2	L'umano come custode dell'etica e della sicurezza	19
6.3	L'evoluzione del ruolo dello sviluppatore	20
6.4	Problemi nel testing automatico tra AI	22
7	L'impatto degli LLM - Opportunità e rischi per gli sviluppatori	23
7.1	Machine Unlearning: Come un'AI disimpara	25
7.2	Modelli senza dati: Un futuro possibile?	26
7.2.1	Neuro-Symbolic AI	28
8	L'AGI (Artificial General Intelligence)	29
9	Perché fermare l'AI non ha senso	33

10 Economia e Disruptive Marketing	34
11 Il futuro dello sviluppatore: Creatore non solo orchestratore	37
11.1 Tra 10 anni: sviluppatore progettista di strumenti AI	38
11.2 Tra 50 anni: sviluppatore come creatore di soluzioni per un mondo inclusivo e sostenibile	39
11.3 Tra 100 anni: sviluppatore come custode dell'umanità digitale	41
11.4 È più facile prevedere ciò che (forse) non accadrà	42
12 Conclusioni	43
12.1 L'esperimento finale	44
12.1.1 Poniamo la stessa domanda a ChatGPT	45
12.2 Considerazioni finali	47

Sommario

In questa relazione saranno, osservate temi inerenti alla natura dell'intelligenza artificiale, con particolare attenzione ai Large Language Models (LLM) e al loro impatto sul ruolo degli sviluppatori, andando ad analizzare le sfide che emergono dall'integrazione di queste tecnologie avanzate nel processo dello sviluppo software.

Alla fine di questa relazione, ci sarà un piccolo esperimento che potrà essere svolto solo dopo aver letto l'intera relazione.

1 Introduzione

All'inizio dell'informatica, figure come Alan Turing (che ad oggi viene considerato il padre dell'informatica e dell'intelligenza artificiale) e altri pionieri gettavano le prime basi per ciò che oggi conosciamo come intelligenza artificiale (AI). Già allora, si interrogavano sulla possibilità di costruire macchine capaci di emulare o addirittura superare il comportamento del cervello umano. In questi ultimi anni, l'intelligenza artificiale (AI) ha compiuto passi da giganti, trasformando il settore tecnologico, dando origine ad un'era pre-LLM e post-LLM. In particolare, i Large Language Models (LLM), come il più famoso chatGPT di OpenAI, oppure come tanti altri LLM esistenti, hanno dimostrato grandi capacità nella generazione di testo, nella comprensione del linguaggio naturale (ad oggi riscontrano ancora difficoltà in lingue come l'arabo, cinese o giapponese) e nell'automazione di compiti complessi. Questa crescita tecnologica esponenziale, ha sollevato tanti dubbi sul futuro del lavoro umano, specialmente nel campo dello sviluppo software. La domanda centrale che guida questa relazione è: **"Gli sviluppatori saranno sostituiti dall'AI?"**. Il successo degli LLM ha acceso la speranza di poter raggiungere l'Artificial General Intelligence (AGI). Molte persone pensano che aumentando i dati di addestramento e la potenza computazionale, potremmo avvicinarci ad un'intelligenza comparabile a quella umana. Tuttavia, reputo che non si potrà raggiungere l'intelligenza umana, ma ci sarà una distinzione tra AI debole ed AI forte.

1.1 Come nasce questa relazione

L'idea di questa relazione nasce durante una lezione dedicata agli LLM, dove il docente pose una domanda molto provocatoria e invitò gli studenti ad alzare la mano se fossero d'accordo, la domanda in questione era: **"L'Intelligenza Artificiale eliminerà il lavoro del programmatore?"**. A quella domanda, molte mani si sono alzate, anzi la maggioranza degli studenti presenti in quella lezione aveva la mano alzata. In quel momento, ho provato una profonda sensazione di paura e incertezza, tanto da mettere in discussione la mia scelta

intrapresa riguardo al percorso di studi in informatica. Vedere così tante mani alzate accentuò i miei dubbi sul mio futuro professionale. Tuttavia, invece di lasciarmi sopraffare da queste emozioni negative, ho deciso di trasformarle in uno stimolo per approfondire realmente cosa siano gli LLM e quali siano le loro implicazioni. Questa riflessione mi ha portato a scrivere questa relazione con l'obiettivo di esplorare le capacità e i limiti degli LLM, affrontare le mie paure e riaffermare la mia fiducia nelle mie competenze e nella mia scelta di carriera.

Prima di approfondire questa tematica, voglio presentare un'immagine che richiama quella mostrata dal docente alla fine della lezione:

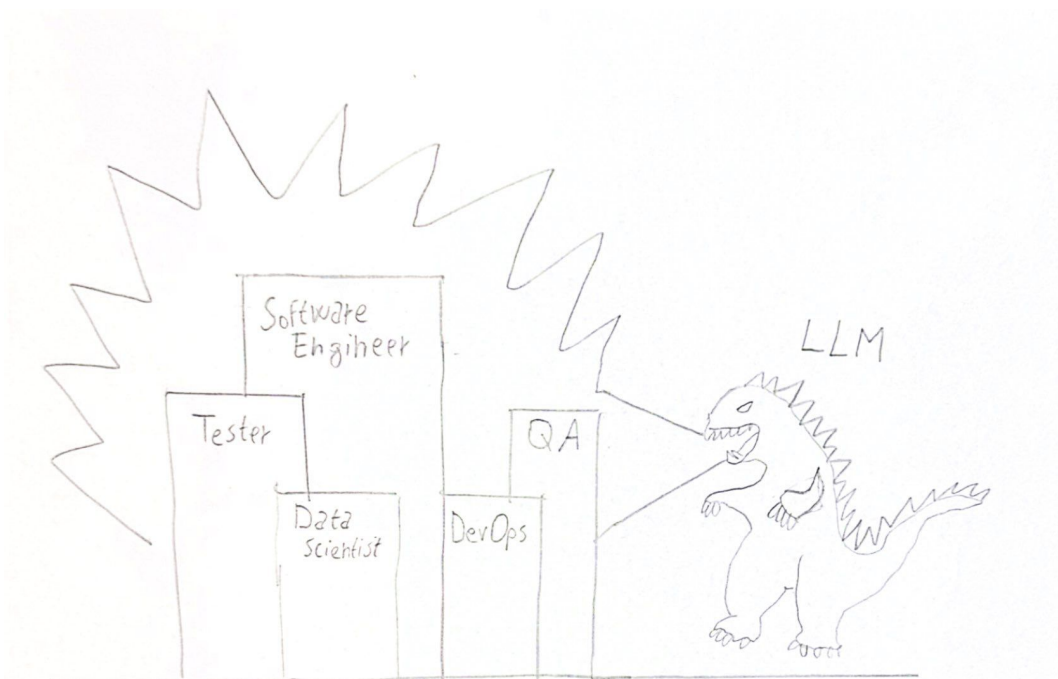


Figura 1: Immagine mostrata durante l'ultima lezione dedicata agli LLM

L'immagine ritrae Godzilla (che rappresenta gli LLM), che distrugge le torri (che rappresentano le figure come sviluppatori, tester, QA, data scientist, ed altre figure legate al settore informatico), la metafora descrive l'impatto potenziale degli LLM sulle professioni legate al mondo dell'informatica. Sebbene la metafora sia efficace per attirare l'attenzione, questa immagine appare

eccessivamente allarmista e non del tutto rappresentativa del reale impatto di queste tecnologie sul mondo del lavoro e sull'evoluzione delle professioni. L'obiettivo di questa relazione è dimostrare che l'immagine di Godzilla che distrugge le torri delle figure legate al mondo dell'informatica è una metafora utile per stimolare la riflessione, ma non riflette la realtà. Gli LLM non sono strumenti destinati a sostituire completamente gli sviluppatori, bensì a potenziarne le capacità, supportando e migliorando il lavoro umano, affiancandosi al loro operato anziché sostituirlo completamente, personalmente considero l'AI come il perfetto stagista, ossia, bravissimo ma da accompagnare e controllare. Questa relazione intende ribaltare la metafora del "Godzilla distruttore" per proporre quella di una convivenza equilibrata, in cui l'AI diventa un prezioso alleato nell'evoluzione professionale e personale dell'essere umano. L'obiettivo è dimostrare che l'intelligenza artificiale non distrugge, ma trasforma e arricchisce il ruolo dello sviluppatore, offrendo nuove opportunità di crescita e innovazione.

1.2 Argomenti della relazione

Le seguenti sezioni si muoveranno lungo più direzioni. In primo luogo, sarà chiarita la natura del pensiero umano e la sua irriducibilità a un modello probabilistico come quello degli LLM, dimostrando che parlare di macchine pensanti è, per ora, un'ipotesi non corretta. In secondo luogo, si analizzerà la cosiddetta bugia degli LLM, cioè il loro apparente potere creativo, che in realtà si basa su meccanismi di ricombinazione statistica anziché su intuizione, emotività ed esperienze personali. Da lì, ci si avvicinerà alla parte centrale, ossia, confutare l'idea che l'AI possa davvero distruggere la figura dello sviluppatore. Al contrario, si mostrerà come il ruolo del programmatore sta evolvendo. Lo sviluppatore non è destinato a scomparire, ma a cambiare, rafforzando sempre più le sue competenze critiche, creative e relazionali. Nel frattempo, è fondamentale non confondere la potenza statistica degli LLM con il concetto di pensiero, le macchine, ancora oggi, non possiedono intenzionalità, coscienza, moralità ed etica. Un conto è generare testo in modo plausibile, un altro è

vivere le esperienze, provare emozioni, compiere scelte etiche, operare con una visione che trascenda la somma dei dati di addestramento. Parliamo di come spesso il marketing influenzi la nostra percezione. Infine, la relazione proporrà un percorso di riflessione sul futuro. Spiegheremo come fermare l'AI non abbia senso, in ragione del fatto che le tecnologie, una volta giunte a un determinato livello di maturità, trovano comunque il modo di diffondersi e di radicarsi. La vera sfida non è ostacolare l'innovazione, bensì governarla, incanalandola in percorsi di sviluppo sostenibile, etico e consapevole. Infine ci saranno con le considerazioni finali.

2 Cos'è il pensiero? L'umano vs la macchina

Nel parlare di LLM, si fa spesso confusione tra pensiero e generazione di testo ammissibile. È fondamentale chiarire, in modo dettagliato, cosa significhi realmente pensare e perché il cervello umano non possa essere ridotto a un semplice sistema di input-output come un computer, per quanto possa essere complesso. La questione del pensiero umano trova le sue radici in millenni di filosofia, a partire dai più famosi filosofi, da Platone, Aristotele, passando per Cartesio, fino a Kant, Popper e arrivando ai contemporanei. In parallelo, negli ultimi decenni, la scienza cognitiva e le neuroscienze hanno gettato nuova luce sulla natura della coscienza, dell'elaborazione delle informazioni da parte del cervello e dell'emergere di abilità come creatività, intuizione, empatia ed emozione. C'è da sottolineare che il modo in cui si origina un pensiero è tuttora ignoto, pertanto nelle neuroscienze il cervello è considerato il mezzo tramite il quale i pensieri vengono resi possibili. Anche la coscienza è un campo ignoto per la neuroscienza, essa può essere definita come l'insieme delle esperienze soggettive e della consapevolezza di sé e del mondo circostante. Essa comprende la capacità di percepire, pensare, sentire e riflettere sulle proprie esperienze. La coscienza umana è strettamente legata alla nostra identità, alle nostre emozioni e alla nostra capacità di creare significati e valori. Le macchine, sono sistemi di calcolo che possono essere estremamente potenti nel risolvere problemi ben definiti, come dimostrato da Deep Blue (era un computer prodotto dall'IBM,

progettato per giocare a scacchi, reso famoso vincendo una partita a scacchi contro un campione del mondo in carica, Garry Kasparov il 10 febbraio 1996), da AlphaGo (in grado di sconfiggere il campione mondiale di Go) o dagli stessi LLM capaci di generare testi in modo sofisticato.

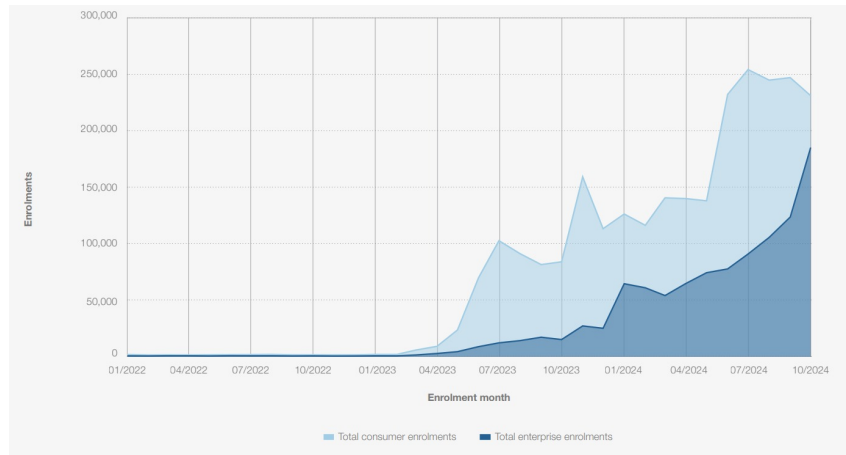


Figura 2: L'aumento dell'interesse per l'AI generativa, evidenziato dalle iscrizioni ai corsi, riflette il crescente dibattito pubblico sulla capacità delle macchine di simulare il pensiero umano. Fonte: World Economic Forum.

Tuttavia, per comprendere la differenza tra il cervello biologico e un modello statistico, dobbiamo definire alcuni concetti chiave:

- **Complessità sinaptica:** Il cervello umano è costituito da circa cento miliardi di neuroni, ciascuno con migliaia di connessioni sinaptiche che cambiano dinamicamente nel corso dell'esistenza. Questa rete di connessioni non è semplicemente programmata, ma si organizza automaticamente attraverso processi evolutivi, genetici e di apprendimento dato dalle esperienze.
- **Emozioni e motivazioni:** Ogni ragionamento umano è influenzato dallo stato emotivo e dall'insieme delle motivazioni personali. Si tratta di fattori che non sono riducibili a semplici funzioni di costo o di probabilità. Le macchine, finora, non possiedono stati emotivi autentici, ma piuttosto possono simulare e riconoscere reazioni emotive se addestrate su certi pattern.

- **Contesto sociale e culturale:** Gli esseri umani sono inseriti in un contesto storico, sociale, linguistico e culturale che ne forma il pensiero. Quando un cervello umano produce un'idea, essa non è soltanto il frutto di un'elaborazione neurale, ma anche di un vissuto che comprende interazioni con altre persone.

2.1 La natura complessa del pensiero umano

Come detto in precedenza, il pensiero umano è frutto di un complesso insieme di processi cognitivi, emotivi e culturali. La mente umana elabora le informazioni integrando componenti razionali (logico deduttive) ed emotive (empatia, coscienza), filtrate da contesti sociali e culturali. L'intuizione è spesso definita come la capacità di cogliere relazioni o soluzioni in maniera non lineare, senza un esplicito ragionamento passo per passo. Le emozioni, influenzano l'interpretazione della realtà e la focalizzazione dell'attenzione, contribuendo a generare soluzioni creative. Ciò che appare straordinariamente affascinante è che non esiste una teoria unificata e definitiva del funzionamento della mente, le neuroscienze continuano a scoprire nuovi aspetti, le teorie della coscienza si susseguono e nessuna sembra riuscire a cogliere totalmente il fenomeno dell'esperienza soggettiva. Questo alone di mistero è, in un certo senso, ciò che rende unica la creatività umana. Purtroppo, anzi per fortuna, il cervello umano non funziona come un computer, il nostro sistema nervoso opera in modo distribuito, dinamico e fortemente interconnesso, con meccanismi di retroazione continui, l'adattabilità del sistema sinaptico e l'emergere di stati di coscienza rendono il cervello molto diverso da un classico computer che comprende codice binario.

2.2 Perché le macchine non possono pensare

Gli LLM (es. ChatGPT, Gemini, LLaMA, PaLM), operano su basi probabilistiche e logiche. Non hanno un corpo, non sperimentano fame, sete, dolore o piacere e non conoscono la paura della morte. Tutto ciò rende il loro pensiero, se così possiamo chiamarlo, profondamente diverso da quello umano, anzi,

secondo molti studiosi e secondo la mia opinione, non è affatto pensiero, essi predicono la probabilità che una certa parola (o token) segua un'altra, sulla base di regolarità e pattern appresi, non è altro che un'elaborazione statistica. Lo stato attuale della ricerca scientifica riconosce che, pur avendo raggiunto livelli alti di potenza e versatilità, le reti neurali non sono dotate di intenzionalità, coscienza o comprensione del significato di ciò che generano, un tema trattato anche dal filosofo John Searle con l'esperimento della stanza cinese, ideato come controesempio rispetto alla teoria dell'intelligenza artificiale forte. Alla base del ragionamento di Searle vi è l'idea che la sintassi non sia condizione sufficiente per la determinazione della semantica. Quando un LLM produce un testo, non lo fa perché vuole o sente che sia giusto o sbagliato, ma semplicemente perché, in base alle sue gigantissime tabelle di probabilità, semmai quella combinazione di termini è statisticamente adatta al contesto del prompt che gli è stato fornito. C'è da ricordare che gli esseri umani sono creatori che infondono significato e intenzione nelle loro opere, siano esse artistiche, scientifiche o tecnologiche. La creazione umana è guidata da esperienze personali, emozioni e obiettivi che riflettono una comprensione profonda del mondo e dei bisogni degli altri. Le macchine, invece, sono simulatori che possono imitare comportamenti umani basandosi su dati e algoritmi, ma senza una comprensione autentica o un'intenzione intrinseca. Dunque il pensiero umano è un fenomeno complesso, integrato con l'emotività, la corporeità, la società e la cultura, mentre gli LLM restano sistemi statistici sofisticati, privi di consapevolezza e di intenzionalità.

3 La filosofia del pensiero e della creazione

Il pensiero e la creazione artistica o intellettuale costituiscono i pilastri dell'essenza umana. Se fino a pochi decenni fa la domanda "le macchine potranno mai pensare?" era di interesse di filosofi e scienziati, oggi, con l'avvento di sistemi come gli LLM, la questione è divenuta di dominio pubblico, alimentando un vivace dibattito. Eppure, per chi si occupa di scienza o di filosofia della mente, è evidente come la separazione fra imitazione e autenticità sia tutt'altro che

risolta. Nei secoli, la filosofia ha approfondito il tema della coscienza e della mente, spesso affrontando quesiti quali la natura del libero arbitrio, l'origine dell'autoconsapevolezza e la differenza tra un'azione meccanica e un atto intenzionale. Se consideriamo la mente umana come un sistema emergente, dove miliardi di neuroni interagiscono e danno vita alla soggettività, è difficile immaginare che un computer, per quanto vasto, possa replicare appieno i fenomeni qualitativi, cioè le esperienze soggettive e i *qualia* (i qualia sono le esperienze soggettive e fenomeniche della percezione, come il colore rosso di una mela o la sensazione di calore al sole, ossia, quegli elementi che non possono essere spiegati oggettivamente). Gli LLM, tuttavia, hanno raggiunto un livello tale da ingannare i non addetti ai lavori, molto spesso anche con chi ci lavora, che vedono in essi una vera e propria intelligenza paragonabile all'umana. Parte della confusione nasce dal cosiddetto effetto Eliza (un fenomeno psicologico che porta le persone ad attribuire qualità umane a sistemi informatici che rispondono in modo apparentemente sensato, prende il nome dal programma ELIZA, sviluppato negli anni '60 da Joseph Weizenbaum, che simulava una conversazione con uno psicoterapeuta). Sebbene quel programma seguisse schemi linguistici semplici, molti utenti percepivano una comprensione autentica da parte della macchina. Tuttavia, risposte coerenti non significano comprensione, così come una poesia, canzone o un testo generato da un LLM non implica un'espressione di vissuto interiore. In filosofia, si distinguono tra problemi facili e problemi difficili della coscienza, i primi riguardano i meccanismi di elaborazione dell'informazione, mentre i secondi toccano la sfera soggettiva, la consapevolezza e il vissuto. Gli LLM possono forse avvicinarsi alla risoluzione dei problemi facili (elaborazione di dati e riconoscimento di pattern), ma restano completamente estranei ai problemi difficili, come la comprensione dell'esperienza soggettiva e della coscienza. La motivazione del perché la creazione è un processo umano e non automatizzabile è che essa implica una volontà intrinseca di dare forma a qualcosa di nuovo sulla base dell'esperienza personale, del contesto e delle emozioni. Un pittore non si limita a mescolare colori, piuttosto desidera esprimere una visione del mondo, una sensazione provata in un determinato momento, o mandare un messaggio sociale. Un compositore di musica non

si affida unicamente a combinazioni casuali di note, ma cerca di trasmettere un'emozione o una narrazione. La creazione artistica, così come quella scientifica, implica passione, curiosità, intuito, capacità di visione, elementi che si trovano nell'esperienza umana. Quando un Large Language Model genera un testo, non è guidato da sentimenti o esperienze dirette. La sua produzione si basa su reti neurali addestrate su enormi quantità di dati, che selezionano la sequenza di parole più probabile e coerente rispetto a un determinato prompt. Sebbene una rete neurale possa combinare dati preesistenti e generare contenuti originali nella forma, manca completamente di una ricerca di senso o di una spinta intenzionale a orientare il processo creativo. Di conseguenza, ciò che produce non può essere considerato una creazione autentica, poiché è privo di intenzionalità e significato profondo. Oggi esistono sistemi di intelligenza artificiale come DALL-E e Midjourney che possono produrre opere pittoriche nello stile di un determinato artista. Queste opere possono essere esteticamente sorprendenti, tecnicamente complesse, ma il sistema non percepisce alcuna emozione, non ha riflettuto sul senso dell'immagine, non ha scelto volontariamente di rappresentare un sentimento d'angoscia o di gioia. È un lavoro di modellazione statistica che, per quanto sofisticato, non implica un coinvolgimento emotivo. Le macchine possono fornire un supporto rilevante nella risoluzione di problemi, ma non decidono da sole che un problema meriti di essere risolto per un fine etico o morale. Quel tipo di valutazione, quel "cosa è giusto o sbagliato fare", rientra nella sfera dell'etica umana, che è inscindibile dalla comunità, dalla cultura e dalle istituzioni sociali. Ecco perché, a un livello più avanzato, quando si afferma che l'AI distruggerà lo sviluppatore, si trascura il fatto che il processo stesso di sviluppo del software, nella sua interezza, non è riducibile a un semplice output generato da un modello. Vero è che la scrittura di porzioni di codice potrà essere automatizzata, ma la definizione dei requisiti, l'analisi delle implicazioni etiche, l'allineamento del progetto con la strategia di un'azienda o con il bene di una comunità umana, restano mansioni di natura creativa, che appartengono alla sfera umana.

3.1 Dove gli LLM falliscono nel creare qualcosa di nuovo

Per comprendere appieno i limiti creativi degli LLM, è utile conoscere alcuni esempi pratici in cui queste tecnologie falliscono nel generare contenuti veramente innovativi o significativi, di seguito elencherò soltanto qualche esempio più significativo:

- **Progetti di ricerca originale:** La ricerca scientifica non si limita alla semplice aggregazione e analisi di dati esistenti. Essa richiede la formulazione di ipotesi innovative, la progettazione di esperimenti originali e la capacità di interpretare risultati in contesti complessi e dinamici. Un LLM può assistere nella revisione della letteratura o nella generazione di ipotesi basate su dati preesistenti, ma non può condurre esperimenti fisici, osservare fenomeni o sviluppare nuove teorie scientifiche senza la guida e l'intuizione umana.
- **Innovazioni tecnologiche:** L'innovazione nel campo tecnologico spesso nasce da una combinazione di intuizione, esperienza pratica e una profonda comprensione delle esigenze umane. Un LLM può suggerire miglioramenti o ottimizzazioni basate su dati storici, ma non può intuire esigenze emergenti o sviluppare soluzioni completamente nuove che non siano già state esplorate nei dati di addestramento.
- **Opere artistiche:** Nel campo dell'arte, la capacità di esprimere emozioni profonde e di trasmettere significati complessi è essenziale. Un LLM può generare testi poetici o narrative che rispettano le regole stilistiche di un determinato genere, ma manca della capacità di provare le emozioni che danno vita a quelle opere. Di conseguenza, le creazioni generate dall'AI possono risultare tecnicamente corrette ma prive di autenticità emotiva, mancando di quelle caratteristiche interiori che caratterizza le opere umane.

La consapevolezza dei limiti creativi degli LLM ha importanti implicazioni per il loro utilizzo in ambiti che richiedono innovazione e originalità. Affidarsi completamente a questi modelli per compiti creativi può portare a risultati

prevedibili e poco ispirati, che mancano della profondità e della significatività che caratterizzano le creazioni umane. Inoltre, l'uso improprio degli LLM in contesti creativi può contribuire a una standardizzazione dei contenuti, riducendo la diversità e la ricchezza delle espressioni culturali. Per evitare tali problematiche, è essenziale integrare gli LLM come strumenti di supporto piuttosto che come sostituti dei creatori umani. Gli LLM possono essere utilizzati per generare idee preliminari, per fornire spunti o per assistere nella stesura di contenuti, ma il processo creativo deve rimanere guidato dall'intuizione, dall'esperienza e dalla sensibilità umana.

4 Principali rischi legati agli LLM

Una volta chiarito che gli LLM non hanno coscienza né comprensione autentica, sorge spontanea la domanda, come possono allora diventare pericolosi? Se non pensano, come potrebbero danneggiare gli esseri umani? La risposta a queste domande si trova proprio nella potenza che questi modelli statistici hanno nell'elaborare e generare quantità immense di contenuti, unita all'assenza di qualsiasi filtro interno di tipo etico o intenzionale.

4.1 Rischi nell'uso e nell'implementazione degli LLM

Un LLM non sceglie autonomamente di creare un contenuto malevolo o discriminatorio. Tuttavia, se viene addestrato su dati che contengono bias o pregiudizi, o se viene istruito da un utente malintenzionato a produrre materiale dannoso (testi manipolativi, propaganda, phishing), può farlo con grande efficienza e velocità, amplificando esponenzialmente l'impatto di un comportamento che nasce dalla volontà umana. Senza un elemento di supervisione umana e senza un contesto etico, l'LLM non è in grado di decidere cosa sia lecito o illecito diffondere. C'è anche il rischio che l'impiego indiscriminato di sistemi di intelligenza artificiale conduca a bias sistematici se i dataset di addestramento contengono pregiudizi. Ne sono esempi eclatanti i casi documentati di discriminazione nel riconoscimento facciale e nelle valutazioni del

credito. Un caso significativo è quello di Amazon, che aveva sviluppato un sistema di selezione dei CV che penalizzava automaticamente le candidature femminili, escludendo le donne da posizioni tecniche e di leadership. Un altro caso riguarda un sistema di riconoscimento delle immagini di Google Photos, che classificava in modo offensivo alcune persone di colore come gorilla, evidenziando gravi distorsioni nei dati di addestramento. Senza la supervisione di un essere umano, questi bias rischiano di radicalizzarsi, con conseguenze gravi e difficili da correggere. C'è da considerare anche il problema delle allucinazioni, ovvero risposte inventate e non basate sui dati reali generate dagli LLM. Vediamo, dunque, alcuni dei principali problemi di sicurezza e rischi associati all'uso imponente degli LLM, soprattutto in scenari in cui una spinta automazione potrebbe sfuggire a un adeguato controllo. Di seguito sono elencati alcuni aspetti cruciali:

- **La generazione di contenuti dannosi:** Un LLM, su comando, può produrre testi che disinformano, che incitano all'odio, che portano alla truffa o che contengono istruzioni per compiere attività illegali. Se consideriamo l'elevato grado di realismo che tali modelli possono raggiungere, per esempio nella simulazione di identità o stili di scrittura, la possibilità di condurre campagne di phishing o di cattiva propaganda su larga scala si moltiplica. IL malintenzionato può generare email e messaggi particolarmente convincenti, adattati al contesto della vittima.
- **La manipolazione delle informazioni:** Gli LLM potrebbero essere usati per produrre contenuti falsi, diffondendoli sui social media oppure sui siti di informazione, alterando la percezione degli eventi. In questo senso, la potenza generativa di un LLM può diventare un'arma nella cosiddetta guerra dell'informazione, in cui le notizie, vere o false, vengono veicolate in modo sistematico per influenzare l'opinione pubblica, destabilizzare istituzioni o compromettere la fiducia nelle fonti ufficiali.
- **Presenza di dati sensibili nei dataset di addestramento:** Se l'LLM è stato addestrato su dati che contengono informazioni riservate, potrebbe in certe condizioni rilevare frammenti di questi dati nelle sue rispo-

ste. È un problema di sicurezza e di privacy rilevante, che rischia di violare normative di trattamento dei dati, come per esempio il Regolamento Generale sulla Protezione dei Dati (GDPR) dell'Unione Europea. La protezione dei dati personali diventa cruciale in un contesto in cui tali informazioni sono gestite da sistemi che non possiedono una reale comprensione o capacità decisionale etica. Inoltre, i modelli possono essere esposti a tentativi di manipolazione intenzionale da parte di attaccanti. Uno degli attacchi più comuni è il prompt injection, in cui un utente malintenzionato inserisce nel testo di input istruzioni ambigue o ingannevoli per indurre il modello a produrre contenuti non desiderati, rivelare informazioni riservate o persino generare output potenzialmente pericolosi.

- **La black box dei modelli:** Gli LLM funzionano come una black box in cui è difficile capire come una certa risposta venga costruita a partire da un particolare input. Questa mancanza di trasparenza rende più complicata l'analisi dei rischi e la garanzia di conformità a standard e normative. Ciò rende difficile attribuire responsabilità in caso di errori.

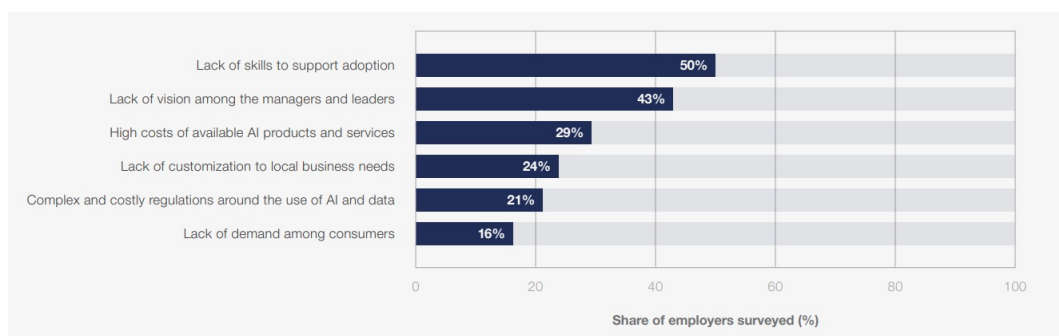


Figura 3: Le barriere all'adozione dell'AI, come la mancanza di competenze e le sfide etiche, mettono in evidenza i rischi di un'implementazione poco consapevole degli LLM. Fonte: World Economic Forum.

L'adozione diffusa dell'intelligenza artificiale ha sollevato una serie di questioni etiche, regolamentari e ambientali che richiedono un intervento coordinato a livello globale. La mancanza di un quadro normativo uniforme può portare a disparità nell'uso e nell'implementazione dell'AI, con gravi conseguenze per la

privacy, la sicurezza e i diritti umani. Organizzazioni internazionali come l'Unione Europea, l'Organizzazione per la Cooperazione e lo Sviluppo Economico (OCSE) e l'Unesco hanno avviato iniziative per sviluppare linee guida etiche che stabiliscano principi fondamentali, quali trasparenza, equità, responsabilità e sicurezza. Questi principi devono essere integrati in ogni fase dello sviluppo e dell'implementazione delle tecnologie AI. Non possiamo, però, ignorare i problemi ambientali legati alla crescita dell'AI, che rappresentano un problema importante e spesso sottovalutato. L'addestramento e il funzionamento dei modelli di intelligenza artificiale, in particolare i modelli di grandi dimensioni, richiedono una quantità significativa di risorse computazionali ed energia. Questo si traduce in un impatto ambientale elevato, con emissioni di carbonio che contribuiscono in maniera negativa al cambiamento climatico. Secondo uno studio dell'università del Massachusetts, negli Stati Uniti, che l'addestramento di un singolo modello AI può generare emissioni di 284 tonnellate di anidride carbonica. Cinque volte l'impatto ambientale medio di un'automobile durante il suo intero ciclo di vita. Inoltre, l'infrastruttura necessaria per supportare l'AI come i data center, server e dispositivi connessi, richiede un consumo energetico costante e intensivo. La crescente domanda di risorse computazionali, se non adeguatamente gestita, rischia di aggravare ulteriormente la crisi ambientale. Per affrontare questa problematica, è fondamentale adottare strategie volte a ridurre l'impronta ecologica dell'intelligenza artificiale. Queste strategie includono l'ottimizzazione degli algoritmi per ridurre il consumo energetico, l'adozione di data center alimentati da fonti rinnovabili e lo sviluppo di modelli più piccoli e efficienti dal punto di vista computazionale. Per non dilungarmi troppo, l'intelligenza artificiale offre enormi opportunità, ma il suo sviluppo e la sua diffusione devono essere guidati da un approccio che consideri non solo le implicazioni etiche e regolamentari, ma anche l'impatto ambientale. Solo in questo modo possiamo garantire un futuro in cui la tecnologia sia al servizio dell'umanità e alla sostenibilità del pianeta. Gli LLM sono in definitiva uno strumento potentissimo, ma privo di moralità, intenzione o comprensione dell'impatto delle proprie azioni. Pertanto, diventa cruciale il ruolo di chi lo progetta, lo addestra e lo utilizza. È qui che, ancora una vol-

ta, la figura dello sviluppatore (insieme agli esperti di regolamentazione e ai responsabili etici) si rivela insostituibile. Qualcuno deve assumersi il compito di definire policy di moderazione dei contenuti, di filtrare i prompt malevoli, di impostare soglie di allarme nel caso in cui l'output dell'LLM assuma toni offensivi o manipolativi e dare un occhio alla sostenibilità ambientale, anzi, due occhi. Questa riflessione sui problemi principali degli LLM sulla sicurezza e sull'ambiente non è fine a sé stessa. Serve a spiegare perché, gli LLM non possano distruggere la figura umana che si trova dietro al software. L'avanzata degli LLM rende ancora più necessaria la presenza di professionisti che abbiano competenze miste, da un lato, la comprensione approfondita della tecnologia AI, dall'altro lato, la sensibilità etica e legale per anticipare gli scenari di rischio e gestirne le conseguenze. La supervisione continua, la definizione di standard e linee guida, la capacità di educare gli utenti e di intervenire quando qualcosa va storto sono tutti aspetti che richiedono capacità umane di giudizio, esperienza e responsabilità.

5 L'etica nell'AI

L'etica è una componente cruciale nello sviluppo e nell'applicazione dell'intelligenza artificiale, per comprenderne l'importanza, possiamo partire da un famoso esperimento filosofico, il *trolley problem*. Immaginate un vagone senza controllo che sta per investire cinque persone legate al binario. Una leva, se tirata, potrebbe deviare il vagone su un altro binario, sacrificando però una persona. Tu sei l'unico vicino alla leva. Che cosa faresti? Sacrificherei una persona per salvarne cinque o preferiresti non agire? Questo problema, formulato dalla filosofa Philippa Foot nel 1967, illustra la difficoltà di prendere decisioni etiche. A complicare la questione, possiamo introdurre ulteriori variabili, e se quella persona fosse qualcuno a te caro? E ancora, se al posto della leva ci fosse un sistema di intelligenza artificiale, quale decisione dovrebbe prendere?

5.1 Il trolley problem applicato all'AI

Nel contesto dell'intelligenza artificiale, il trolley problem diventa più di un esperimento teorico. Immaginate un'auto a guida autonoma che deve scegliere tra sterzare e colpire una persona sul marciapiede o mantenere la traiettoria e investire cinque persone che attraversano la strada. A quale etica dovremmo insegnare alla macchina di aderire? E soprattutto, come possiamo essere certi che le decisioni prese siano giuste? Il problema si complica ulteriormente con l'aggiunta di variabili personali e sociali. Se l'auto dovesse scegliere tra investire una bambina che attraversa col rosso o un'anziana signora sul marciapiede, quale sarebbe la decisione eticamente corretta? Istintivamente, molti potrebbero scegliere di salvare la bambina, considerando che ha più anni di vita davanti a sé. Ma cosa succederebbe se le persone sul marciapiede fossero più di una? E se l'auto dovesse scegliere tra salvare un criminale o un soccorritore? Questi scenari evidenziano come i bias o pregiudizi umani, pregiudizi consci o inconsci, possano influenzare il nostro giudizio e, di conseguenza, la programmazione etica delle macchine.

5.2 Bias e complessità etica

Le decisioni etiche non sono mai universali, poiché dipendono dal contesto culturale, sociale ed emotivo di ciascun individuo. Per esempio, alcune culture potrebbero dare maggiore importanza alla vita dei giovani, mentre altre potrebbero valorizzare il rispetto per gli anziani. Inoltre, il valore che attribuiamo alla vita di una persona varia in base alla nostra percezione, qualcuno a noi caro sarà sempre considerato più importante rispetto a uno sconosciuto, e una persona percepita come cattiva potrebbe essere considerata meno degna di essere salvata. Questo è il cuore del problema etico per l'AI, come possiamo programmare un sistema per prendere decisioni morali in situazioni complesse? Le macchine, a differenza degli esseri umani, non hanno emozioni, intuizioni o valori personali. Le loro decisioni si basano sui dati con cui sono state addestrate, che però possono riflettere i bias presenti nella società. Ad esempio, un algoritmo potrebbe assegnare un valore diverso a una vita in

base al genere, all'etnia o all'età, tramandando ingiustizie e discriminazioni. Le auto a guida autonoma rappresentano una delle applicazioni più avanzate dell'AI, promettendo di ridurre drasticamente gli incidenti stradali, per lo più causati da errori umani come distrazione, stanchezza o guida sotto l'effetto di alcol. Questi veicoli non si distraggono, rispettano i limiti di velocità e hanno una visione a 360° grazie a telecamere e sensori avanzati. Da un punto di vista tecnologico, le auto a guida autonoma richiedono enormi quantità di dati per funzionare correttamente e dipendono da infrastrutture stradali ben mantenute e mappate. Quando si verificano incidenti, emergono problemi di responsabilità, se un veicolo autonomo causa un danno, chi è da ritenere responsabile? Il passeggero, che non stava guidando? Il produttore del veicolo? O l'ingegnere che ha progettato il software? Inoltre, un errore umano casuale è spesso percepito come più accettabile rispetto a un errore commesso da un software. Questo mette in evidenza l'importanza di costruire fiducia nelle tecnologie autonome e di quanto sia fondamentale ridurre gli errori da parte delle macchine.

5.3 Un'etica è programmabile?

La tecnologia, compresa l'intelligenza artificiale, non può esistere senza implicazioni etiche. È fondamentale definire delle regole sociali comuni che stabiliscano dei principi per guidare lo sviluppo e l'applicazione dell'AI. Questo contratto dovrebbe includere:

- **Trasparenza:** Gli algoritmi devono essere comprensibili e spiegabili, affinché gli utenti possano capire come vengono prese le decisioni.
- **Responsabilità:** Devono essere chiaramente definiti i ruoli e le responsabilità in caso di incidenti o errori.
- **Universalità:** Le decisioni etiche delle macchine devono rispettare principi morali condivisi, evitando discriminazioni o ingiustizie.

Il problema dell'etica nell'intelligenza artificiale non ha soluzioni semplici. In primis, gli stessi esseri umani faticano a raggiungere un consenso su ciò che è

giusto o sbagliato in situazioni morali complesse. Tuttavia, ciò non significa che dobbiamo rinunciare a cercare risposte. L'AI deve essere addestrata per prendere decisioni etiche, basandosi su principi chiari e condivisi, ma dobbiamo anche riconoscere i suoi limiti. Le macchine possono ridurre significativamente gli errori e migliorare la sicurezza, ma non saranno mai infallibili. Come società, il nostro compito è quello di prepararci a convivere con queste tecnologie, comprendendone le potenzialità e accettandone i limiti, consapevoli che le decisioni morali, per fortuna, almeno per ora, spettano ancora a noi.

6 Perché l'AI non sostituisce lo sviluppatore

Arriviamo al cuore della discussione, ossia all'idea che un'intelligenza artificiale avanzata possa sostituire totalmente la figura professionale dello sviluppatore. Come è possibile che un sistema AI apparentemente così brillante e veloce nell'elaborazione di codice, nella creazione di documentazione e nella risoluzione di problemi, non sia in grado di rimpiazzare l'essere umano che scrive software? La risposta va cercata in più punti, che vanno dal merito prettamente tecnico al tema etico e decisionale.

6.1 Gli LLM simulano ma non comprendono

Le decisioni nello sviluppo software non riguardano solo il funzionamento immediato, ma anche le conseguenze a lungo termine per gli utenti. Uno sviluppatore, immerso nella società e portatore di valori ed esperienze condivise, è in grado di valutare le implicazioni etiche del proprio lavoro.

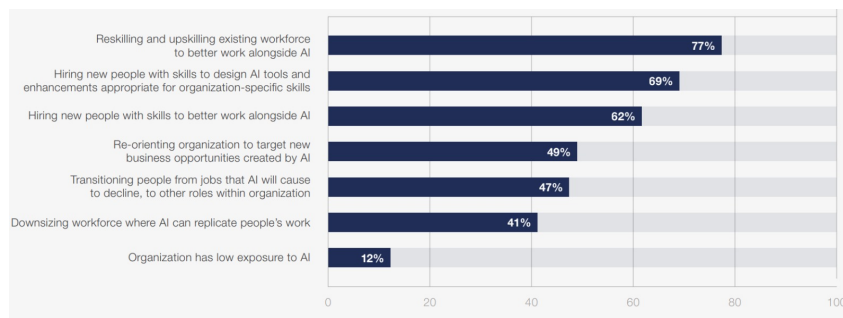


Figura 4: Percentuale di aziende che prevedono di implementare le strategie indicate in risposta alle capacità e alla diffusione crescenti dell'AI. Le strategie aziendali si stanno adattando per integrare l'AI come supporto agli esseri umani, piuttosto che sostituirli, enfatizzando il ruolo insostituibile dello sviluppatore. Fonte: World Economic Forum

Al contrario, un LLM privo di morale ed emotività, non può distinguere tra un software che migliora la vita delle persone e uno che potrebbe danneggiarle. Questo rende evidente come la responsabilità professionale e sociale dello sviluppatore non possa essere delegabile a un algoritmo. Sebbene un LLM sia in grado di generare codice, documentazione e suggerire soluzioni a bug, non comprende il contesto in cui queste soluzioni si applicano né le motivazioni dietro di esse. L'esperienza dello sviluppatore, maturata attraverso la risoluzione di problemi reali, il confronto con i colleghi e il dialogo con i clienti, rimane insostituibile. Affidare completamente lo sviluppo a una macchina statistica, senza supervisione umana, espone a rischi concreti di errori, usi impropri della tecnologia e mancanza di considerazioni etiche. Questo limite diventa ancora più critico in settori sensibili come quello finanziario, medico o aerospaziale, dove le decisioni coinvolgono responsabilità legali ed etiche. Un LLM non prende decisioni, ma fornisce solo risposte probabilistiche. In caso di ambiguità, lo sviluppatore resta sempre il responsabile ultimo delle scelte effettuate, confermando che la supervisione umana è insostituibile per garantire sicurezza, integrità e affidabilità nello sviluppo software.

6.2 L'umano come custode dell'etica e della sicurezza

Lo sviluppatore, ben lontano dall'essere superfluo, diventa più che mai necessario per garantire il rispetto di principi etici e normativi, oltre che per

sorvegliare la sicurezza e la correttezza delle implementazioni. Ad esempio, anche nei processi di testing, l'introduzione di strumenti di AI può velocizzare i controlli iniziali, ma non può sostituire la competenza di chi conosce le specifiche, l'ambiente di esecuzione e le possibili implicazioni di un malfunzionamento. Sostanzialmente, la tesi che l'AI distrugge lo sviluppatore non regge alla prova dei fatti se analizziamo il tema in profondità, ci sono troppe sfumature, troppe responsabilità, troppe variabili che rientrano nel processo di sviluppo per pensare di delegarle completamente a un modello statistico. Anzi, più l'AI si fa potente, più diventa urgente la presenza umana che la disciplini e la orienti.

6.3 L'evoluzione del ruolo dello sviluppatore

Il ruolo dello sviluppatore sta subendo una trasformazione significativa, non si limita più alla scrittura di codice, ma il bacino di competenze diventa più ampio.



Figura 5: Le competenze legate all'AI e ai Big Data sono considerate cruciali dai principali settori industriali, con il settore automobilistico e quello delle telecomunicazioni al primo posto. Fonte: World Economic Forum.

L'introduzione degli LLM e di altre tecnologie di intelligenza artificiale non sostituisce gli sviluppatori umani, piuttosto crea un bisogno crescente di figure capaci di orchestrare sistemi complessi, affrontare sfide etiche e promuovere innovazioni in un contesto tecnologico in continua evoluzione. Tra le compe-

tenze emergenti, una delle più rilevanti è il prompt engineering, che consiste nel formulare richieste efficaci per ottenere risultati ottimali dagli LLM. Questa abilità richiede una profonda comprensione del funzionamento degli LLM. Accanto a questa, gli sviluppatori devono padroneggiare la capacità di integrare metriche di qualità e di validare i contenuti generati dall'AI, garantendo coerenza, precisione e affidabilità dei risultati, oltre a monitorare costantemente le prestazioni dei sistemi.

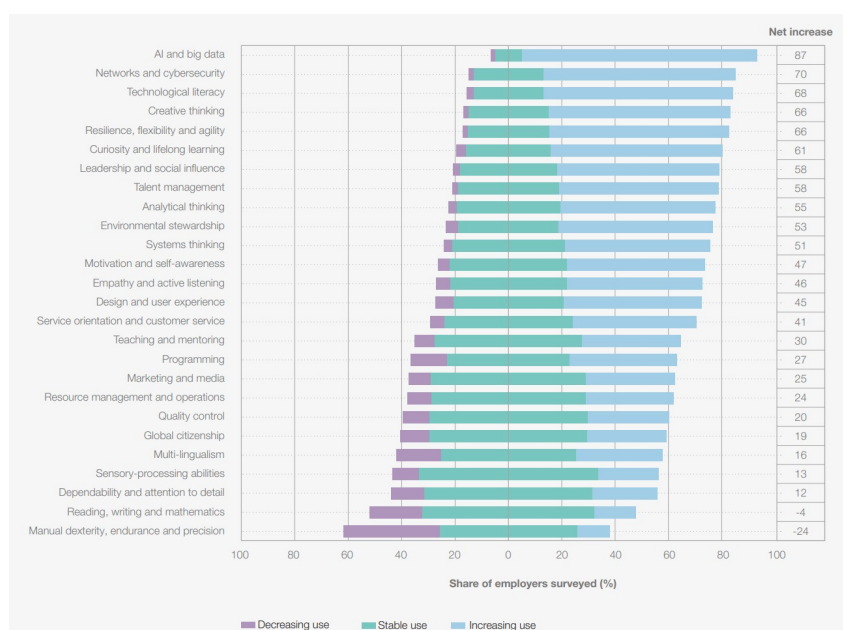


Figura 6: Crescita prevista delle competenze entro il 2030, con un forte aumento per competenze legate all'AI. Fonte: World Economic Forum.

In questa evoluzione, lo sviluppatore diventa un vero e proprio ideatore di soluzioni che vanno oltre l'aspetto tecnico, includendo una visione etica e sociale delle tecnologie. Le nuove responsabilità comprendono non solo la progettazione e la gestione di sistemi AI, ma anche l'ottimizzazione dei prompt e la mitigazione dei rischi legati alla disinformazione e ai bias algoritmici. Lo sviluppatore assume una posizione centrale nel futuro del software, fungendo da ponte tra automazione avanzata e il mondo umano. Questa evoluzione non solo preserva, ma amplifica l'importanza della figura dello sviluppatore.

6.4 Problemi nel testing automatico tra AI

L'automazione del processo di testing tramite sistemi AI che testano altri sistemi AI introduce una serie di rischi e problemi che devono essere attentamente considerati per garantire la qualità e l'affidabilità del software sviluppato. Quando un sistema AI viene utilizzato per testare un altro sistema AI, c'è il rischio che gli errori e i bias presenti nel primo si riflettano e si amplifichino nel secondo. Questo può portare a una propagazione di difetti che è difficile da individuare e correggere senza l'intervento umano. Ad esempio, se un sistema di testing automatizzato ha un bias che lo porta a non rilevare certi tipi di bug, questi bug possono passare inosservati nel software finale, compromettendo la sua qualità e sicurezza. Come detto in precedenza, i sistemi AI sono delle black box in cui è difficile comprendere come vengono prese determinate decisioni o identificati specifici problemi. Questa mancanza di trasparenza rende complicato valutare l'efficacia del processo di testing e identificare le cause degli errori, purtroppo senza una comprensione chiara di come il sistema AI ha effettuato il testing, diventa difficile migliorare e ottimizzare il processo stesso. Inoltre l'utilizzo di sistemi AI che testano altri sistemi AI può creare una dipendenza reciproca, dove la qualità del testing dipende dalla qualità del sistema che effettua il testing. Questo può portare a una catena infinita di sistemi che si affidano a altri sistemi per verificare la loro correttezza. Gli LLM possono avere difficoltà a comprendere il contesto specifico in cui viene utilizzato il software, portando a test che non sono completamente pertinenti o che non coprono tutti i possibili casi d'uso, facendo risultare una copertura di test incompleta, con conseguenti rischi di malfunzionamenti o comportamenti imprevisti del software. Infine affidare totalmente il processo di testing a sistemi AI aumenta anche i rischi di sicurezza. Se un sistema AI di testing viene compromesso, potrebbe introdurre intenzionalmente errori o vulnerabilità nel software sottoposto a testing, compromettendo la sicurezza dell'intero sistema. Dunque, per diminuire i rischi associati al testing automatico tra AI, è essenziale mantenere un ruolo attivo e supervisore degli sviluppatori umani nel processo di testing. Ecco alcune soluzioni possibili:

- **Implementazione di verifiche incrociate:** Gli sviluppatori devono implementare verifiche incrociate utilizzando diversi metodi di testing, combinando sistemi automatizzati e test manuali. Questo approccio aiuta a identificare e correggere gli errori che potrebbero sfuggire ai sistemi di testing automatizzati, garantendo una copertura di test più completa e affidabile. Gli sviluppatori devono controllare e validare i risultati dei test automatizzati. Questo include la revisione dei risultati, l'identificazione di eventuali anomalie e la correzione dei problemi rilevati.
- **Assicurare la trasparenza dei processi decisionali:** È fondamentale sviluppare sistemi di testing AI che siano trasparenti e interpretabili, consentendo agli sviluppatori di comprendere come vengono prese le decisioni e identificati i problemi. L'uso di tecniche di interpretabilità può aiutare a rendere i processi decisionali degli LLM più comprensibili e verificabili.

In questo contesto il ruolo dello sviluppatore come supervisore e garante della qualità del software rimane fondamentale, anche in un contesto dominato dall'automazione e dall'AI. La collaborazione tra uomo e macchina può portare a una maggiore produttività, efficienza e a una migliore qualità del software, ma solo se accompagnata da una supervisione umana attenta e consapevole.

7 L'impatto degli LLM - Opportunità e rischi per gli sviluppatori

L'integrazione sempre più avanzata degli LLM sta modificando profondamente lo sviluppo software, questi strumenti stanno riducendo la necessità di alcune mansioni tradizionali, trasformando il mercato del lavoro nel settore IT. Strumenti come GitHub Copilot, ad esempio, possono generare interi blocchi di codice o funzioni complesse in pochi secondi, accelerando il processo di sviluppo e diminuendo la necessità di scrivere codice manualmente. Allo stesso modo, ChatGPT e modelli di testing automatizzati possono eseguire verifiche

funzionali e di sicurezza senza intervento umano, migliorando l'efficienza e riducendo i tempi di rilascio.

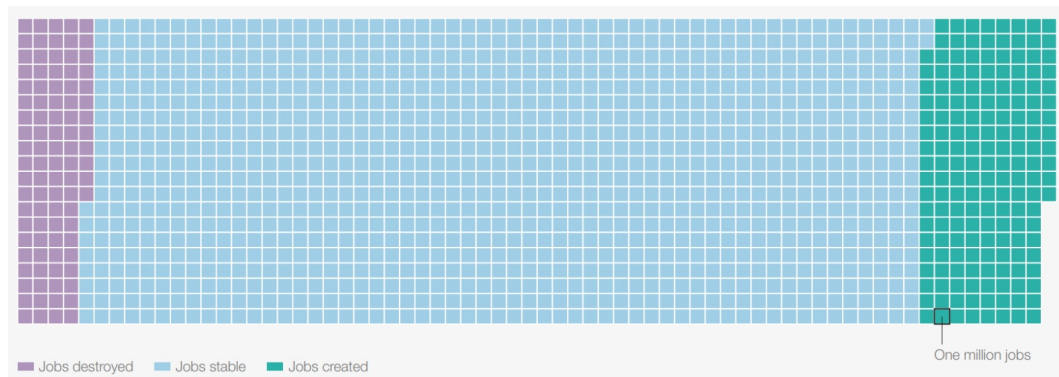


Figura 7: **Proiezione dei cambiamenti globali nell'occupazione entro il 2030:** il bilancio netto di posti di lavoro persi e creati a causa dell'automazione e delle nuove tecnologie. Si stima che nei prossimi cinque anni saranno creati 170 milioni di nuovi posti di lavoro, mentre 92 milioni saranno eliminati, portando a un ricambio strutturale pari al 22% degli attuali posti di lavoro formali. Il risultato netto sarà un aumento del 7% dell'occupazione globale, pari a 78 milioni di posti di lavoro. Fonte: World Economic Forum.

Questa evoluzione, se da un lato promette maggiore velocità e produttività, dall'altro solleva importanti interrogativi. L'automazione basata su LLM potrebbe sostituire alcune figure professionali chiave, come sviluppatori, tester e QA, favorendo una transizione verso un lavoro più semplificato, ma anche meno umano. Le aziende già adottano strumenti come Copilot e ChatGPT per velocizzare i processi, con un impatto significativo sulla struttura e sul funzionamento dei team di sviluppo. Tra le principali implicazioni emerge il rischio di disoccupazione tecnologica. Uno studio recente svolto dall'Università di Trento stima che in Italia, nei prossimi quindici anni, 3,87 milioni di lavoratori saranno a rischio di sostituzione tecnologica nelle singole mansioni, salendo a 7,12 milioni se si considerano le professioni automatizzabili nella loro interezza, a causa dell'automazione e dell'introduzione di tecnologie avanzate come gli LLM. Tuttavia, questo non significa necessariamente una perdita netta di posti di lavoro. Secondo il World Economic Forum, ci saranno più posti creati che persi, nasceranno nuove opportunità lavorative in settori legati alle nuove tecnologie, all'analisi dei dati e alla gestione dei sistemi automatizzati. C'è da dire che gli sviluppatori rischiano di perdere competenze avanzate,

fondamentali per affrontare problemi complessi e innovare creativamente, se si affidano eccessivamente a questi strumenti. Inoltre, la dipendenza da macchine potrebbe compromettere la robustezza dei sistemi, poiché le tecnologie mancano della capacità umana di adattarsi a situazioni impreviste. La creatività e l'innovazione, caratteristiche distintive dell'intelligenza umana, rischiano di essere sacrificate. Gli LLM possono generare contenuti efficaci ma mancano della capacità di formulare idee originali o di contestualizzare le loro soluzioni. Affidarsi completamente a questi strumenti potrebbe portare a una riduzione della diversità e della qualità delle soluzioni software.

7.1 Machine Unlearning: Come un'AI disimpara

Il machine unlearning si riferisce alla capacità di un modello di intelligenza artificiale di rimuovere specifici dati di addestramento su richiesta, senza dover riaddestrare l'intero modello da zero. Questo processo è essenziale per conformarsi alle normative sulla privacy e per garantire che i modelli non mantengano informazioni sensibili che dovrebbero essere cancellate, ma anche nel caso in cui ci siano dati errati all'interno del modello. Questo concetto rappresenta un ostacolo rilevante nell'ambito dell'intelligenza artificiale, soprattutto in relazione al diritto all'oblio (il diritto di un utente a richiedere la rimozione di dati personali dai database delle aziende e delle organizzazioni), definito da normative come il Regolamento Generale sulla Protezione dei Dati (GDPR) dell'Unione Europea. Esistono diverse tecniche per implementare il machine unlearning:

- **Rimozione selettiva dei dati:** Questa tecnica prevede l'identificazione e la rimozione di specifici dati dal set di addestramento, seguita da una rielaborazione dei pesi del modello per riflettere l'assenza di tali dati.
- **Algoritmi di aggiornamento:** Alcuni algoritmi permettono di aggiornare i pesi del modello in modo da dimenticare i dati specifici senza dover riaddestrare l'intero modello. Questi algoritmi modificano i pesi in modo da minimizzare l'influenza dei dati da rimuovere.

- **Mascheramento dei dati:** Questa tecnica prevede l'applicazione di maschere o filtri ai dati da rimuovere, rendendoli inaccessibili al modello senza eliminarli completamente dal set di dati originale.

Implementare il machine unlearning presenta diverse difficoltà tecniche:

- **Rimozione completa dei dati:** Assicurare che i dati specifici siano completamente rimossi dal modello senza lasciare tracce residue, questo perché anche piccoli residui possono compromettere la conformità alle normative sulla privacy.
- **Mantenimento delle performance del modello:** La rimozione di dati specifici può influire negativamente sulle performance del modello. È essenziale bilanciare la necessità di rimuovere i dati con il mantenimento della precisione e dell'efficacia del modello. Inoltre in modelli di grandi dimensioni non è possibile prendere in considerazione l'opzione di riaddestrare l'intero modello per rimuovere dati specifici, questo a causa dei costi computazionali elevati.

Gli sviluppatori hanno un ruolo cruciale nell'implementazione e nella gestione del machine unlearning, poiché devono essere a conoscenza delle leggi e delle normative sulla protezione dei dati che influenzano il machine unlearning, assicurando che le soluzioni implementate siano conformi ai requisiti legali, ed è fondamentale progettare e implementare tecniche di machine unlearning che siano sia efficienti dal punto di vista computazionale che efficaci nella rimozione dei dati specifici e che non influiscano negativamente sulle performance del modello.

7.2 Modelli senza dati: Un futuro possibile?

L'evoluzione degli LLM ha portato allo sviluppo di tecniche avanzate di apprendimento che riducono la dipendenza dai grandi dataset tradizionali. Tre di queste tecniche sono il zero-shot learning, il few-shot learning oppure l'apprendimento per trasferimento, che permettono ai modelli di generalizzare su nuovi compiti con pochissimi o nessun esempio specifico, andando nel dettaglio:

- **Zero-shot learning:** permette ai modelli di affrontare compiti per i quali non sono stati esplicitamente addestrati, basandosi su una comprensione generale del linguaggio e delle relazioni concettuali. Ad esempio, un LLM potrebbe essere in grado di rispondere a domande su un argomento per cui non ha ricevuto dati di addestramento specifici, utilizzando la sua comprensione del linguaggio e delle conoscenze apprese in precedenza.
- **Few-shot learning:** consente ai modelli di adattarsi a nuovi compiti fornendo solo pochi esempi di addestramento. Questo approccio è particolarmente utile in scenari in cui i dati sono limitati o difficili da ottenere. Ad esempio, un modello potrebbe essere in grado di apprendere a tradurre un linguaggio di programmazione con solo pochi esempi di codice tradotto, migliorando la sua capacità di adattarsi a diversi linguaggi di programmazione.
- **L'apprendimento per trasferimento:** consente ai modelli di trasferire conoscenze acquisite in un dominio a un altro dominio correlato. Questo approccio permette di migliorare la flessibilità e la generalizzazione dei modelli, permettendo di applicare le conoscenze apprese in un contesto a nuovi contesti senza la necessità di grandi quantità di dati di addestramento.

Di seguito possiamo elencare alcuni vantaggi e limiti:

- **Vantaggi:**
 1. **Efficienza dei dati:** Le tecniche citate in precedenza riducono significativamente la necessità di grandi dataset, rendendo l'addestramento dei modelli più efficiente e meno costoso.
 2. **Generalizzazione migliorata:** Questi approcci migliorano la capacità dei modelli di generalizzare a nuovi compiti e domini, aumentando la loro flessibilità e adattabilità.
 3. **Rispetto della privacy:** La riduzione della dipendenza dai grandi dataset tradizionali diminuisce i rischi legati alla privacy e alla

sicurezza dei dati, poiché i modelli possono essere addestrati con meno dati sensibili.

- **Limiti:**

1. **Difficoltà nell'apprendimento di concetti astratti:** Nonostante i progressi, l'AI può ancora avere difficoltà a comprendere e generalizzare concetti astratti senza esempi specifici, limitando la sua capacità di affrontare compiti complessi.
2. **Interpretabilità:** I modelli che apprendono in modo simbolico o concettuale sono ancora delle black box, rendendo difficile comprendere come prendono decisioni e quali conoscenze hanno acquisito.

7.2.1 Neuro-Symbolic AI

Un approccio promettente per superare questi limiti è la **Neuro-Symbolic AI**, come descritto dalla pagina Wikipedia "Neuro-Symbolic AI", essa combina le capacità delle reti neurali e dei sistemi simbolici di ragionamento per creare un'intelligenza artificiale robusta, capace di apprendere, ragionare e modellare cognitivamente. Questo approccio mira a superare le limitazioni, mentre le reti neurali eccellono nel riconoscimento di schemi, i sistemi simbolici si distinguono nella deduzione logica, nella pianificazione e nella manipolazione di conoscenze astratte. Secondo gli studiosi Leslie Valiant e Gary Marcus, la costruzione di modelli cognitivi ricchi e affidabili richiede un'integrazione tra ragionamento simbolico e apprendimento automatico. Come afferma Marcus, "Non possiamo costruire modelli cognitivi adeguati senza un'architettura ibrida, ricca conoscenza pregressa e tecniche sofisticate per il ragionamento". L'abilità di rappresentare e manipolare conoscenze astratte tramite simboli è essenziale, poiché molte forme di conoscenza sono troppo astratte per essere gestite unicamente dalle reti neurali. Questa visione è rafforzata dai lavori di Henry Kautz, Francesca Rossi e Bart Selman, che si rifanno alla teoria dei due sistemi cognitivi descritta da Daniel Kahneman nel suo libro "Thinking, Fast and Slow". Il "Sistema 1" è veloce, intuitivo e inconscio, perfetto per il riconoscimento di schemi, mentre il "Sistema 2" è lento, esplicito e deliberativo,

ideale per il ragionamento logico. Nell'AI, il primo tipo di cognizione è ben gestito dalle reti neurali profonde, mentre il secondo richiede il ragionamento simbolico. La loro combinazione consente di creare modelli capaci di apprendere, pianificare e interagire in modo efficace con gli esseri umani. Nonostante il potenziale, l'integrazione di questi due paradigmi rimane di una difficoltà elevata. Lo sviluppo di nuove architetture e algoritmi che possano sfruttare i punti di forza di entrambi i sistemi è fondamentale per realizzare un'intelligenza artificiale affidabile, capace di apprendere, ragionare e rispondere a domande complesse. La Neuro-Symbolic AI, rappresenta un passo molto ambizioso e importante per cercare di affrontare problemi complessi che richiedono sia intuizione che deduzione.

8 L'AGI (Artificial General Intelligence)

Nei capitoli precedenti abbiamo analizzato le ragioni per cui un Large Language Model, pur essendo straordinariamente potente, non può sostituire lo sviluppatore umano, soprattutto a causa della mancanza di comprensione, intenzionalità e capacità di operare scelte etiche o contestualizzate. A questo punto, però, è necessario ampliare la riflessione per includere un tema ancora più ambizioso, la prospettiva, almeno teorica, di sviluppare un'Artificial General Intelligence (AGI), ovvero un'intelligenza artificiale in grado di possedere, o quantomeno emulare in modo completo, le abilità cognitive umane. Questa idea, che per decenni è sembrata caratteristica esclusiva della fantascienza, affonda in realtà le sue radici nelle primissime riflessioni di Alan Turing e di altri pionieri dell'informatica. Già nei primi lavori sulla logica computazionale ci si interrogava sulla possibilità di costruire macchine capaci di comportarsi come un cervello umano, o addirittura di superarlo. Tuttavia, il percorso storico dell'AI ha subito oscillazioni notevoli, a volte si è esagerato con l'ottimismo, predicendo un vicino avvento di macchine senzienti, altre volte si è caduti in un eccessivo pessimismo, etichettando la ricerca come irrealistica e priva di risultati. Il successo degli LLM e di altre forme di deep learning ha riaperto, negli ultimi anni, la speranza di poter compiere un salto qualitati-

vo verso l'AGI. Banalmente, si può pensare che, aumentando ulteriormente i dati di addestramento e la potenza computazionale, ci si avvicinerebbe in modo considerevole a una forma di intelligenza comparabile a quella umana. Personalmente ritengono che l'approccio attuale, basato quasi esclusivamente sull'analisi statistica dei dati, sia intrinsecamente limitato e non consentirà mai di varcare la soglia che separa l'AI debole (quella specializzata in compiti specifici) dall'AI forte (dotata di coscienza e comprensione). Prima di entrare nel merito delle sfide tecniche, è però opportuno chiarire i concetti filosofici e cognitivi sottostanti. Quando si parla di general intelligence, non ci si riferisce soltanto a un potenziale aumento delle performance computazionali. Anzi, i tratti distintivi di un'AGI riguardano proprio la capacità di ragionare in modo di capire il linguaggio in profondità (non limitandosi a una correlazione statistica tra stringhe), di stabilire connessioni e analogie tra domini apparentemente distanti, di porsi obiettivi a lungo termine e di modificarli in base alle circostanze, e, per alcuni, perfino di possedere una forma di introspezione o di coscienza. È qui che la questione diventa complessa, poiché la coscienza rimane uno dei più grandi misteri della filosofia e delle scienze cognitive. Al momento, tuttavia, l'AGI resta un obiettivo altamente speculativo. Nessun modello esistente è in grado di dimostrare la flessibilità cognitiva e la consapevolezza tipiche di un essere umano. Una vera AGI, secondo alcune definizioni, dovrebbe non solo simulare il comportamento di un cervello umano, ma anche sviluppare qualcosa di simile a una vita mentale autonoma, che comprende stati interni e intenzioni non immediatamente determinate dall'esterno. Questo significa che dovremmo assistere, in ipotesi, al sorgere di un io computazionale, con una storia di apprendimento e di esperienze che configurino un'identità, qualcosa di paragonabile all'evoluzione del bambino che cresce e impara a conoscere il mondo. In assenza di tale componente identitaria, si può parlare di sistemi molto evoluti, ma non di una mente in senso proprio. Come si inserisce tutto ciò nel discorso sullo sviluppatore? Da un lato, se si immaginasse realmente un'AGI in grado di eguagliare o superare l'intelligenza umana in tutti i compiti, compresa la scrittura di software o la pianificazione di architetture complesse, allora si porrebbe davvero la possibilità che il ruolo dello sviluppa-

tore venga intaccato drasticamente. Dall'altro lato, lo stesso concetto di AGI implica una serie di interrogativi che vanno al di là della sola dinamica lavorativa, sfociando in questioni etiche, filosofiche e persino metafisiche, chi avrebbe la responsabilità di un'AGI se questa prendesse decisioni autonome? Che ne sarebbe dei diritti di una macchina cosciente, qualora lo diventasse? Come prevenire rischi distruttivi, qualora un'AGI decidesse di perseguire obiettivi in conflitto con quelli umani? Senza dubbio, la prospettiva dell'AGI comporta la necessità di una regolamentazione a livello internazionale e di un ripensamento profondo di molti aspetti della nostra società, compreso il lavoro. Al momento, tuttavia, la realtà è che ciò che chiamiamo Artificial Intelligence è ancora molto distante da una vera Artificial General Intelligence. I sistemi di deep learning attuali, per quanto sorprendenti nelle prestazioni, sono essenzialmente specializzati nel riconoscimento di pattern all'interno di dataset di grandi dimensioni e non hanno la flessibilità e la consapevolezza per agire in scenari non strutturati con la stessa discrezione di un essere umano. Questo non significa che dobbiamo sottovalutare i progressi futuri o la potenzialità di nuove architetture ibride, che integrino reti neurali e ragionamento simbolico, oppure che utilizzino tecniche di apprendimento per rinforzo con un alto grado di autonomia. Magari si potrebbe sperimentare forme di auto-addestramento in cui un agente intelligente fa ipotesi sul mondo simulato e corregge i propri errori in base ai risultati delle proprie azioni. In linea teorica, la strada per l'AGI potrebbe passare attraverso l'evoluzione di tali agenti in ambienti sempre più complessi e realistici. Eppure, anche se un giorno dovessimo assistere alla nascita di un'AGI, non è detto che questa realtà avrebbe come effetto la totale cancellazione dello sviluppatore. In ogni rivoluzione tecnologica, infatti, si è visto come la domanda di figure umane competenti cresca, anziché diminuire, almeno nell'immediato orizzonte. Sarebbero necessarie, per esempio, intere nuove generazioni di ricercatori e ingegneri capaci di capire, progettare e mantenere i sistemi di AGI, affrontando tematiche inedite di sicurezza informatica e gestione etica. Inoltre, se davvero le macchine arrivassero a eguagliare o superare le potenzialità umane, si aprirebbe lo scenario di una collaborazione su progetti e attività di portata globale, in cui la presenza

di un'interfaccia umana rimarrebbe fondamentale per interpretare gli obiettivi della società e preservare i valori essenziali dell'umanità. Un altro aspetto spesso trascurato è la questione dell'evoluzione della domanda di lavoro e del ruolo professionale, che non si limita ai soli aspetti tecnici, ma include anche la gestione dei processi. Gli esseri umani, dotati di empatia, capacità di interazione sociale e abilità di leadership, rimangono insostituibili nell'organizzare e motivare squadre di persone, nell'affrontare problemi relazionali e nel costruire contatti. Una potenziale AGI potrebbe, forse, elaborare strategie perfette e ottimizzate sotto il profilo computazionale, ma non avrebbe una comprensione della costruzione emotiva e culturale all'interno del quale si muove la società umana, a meno di non supporre che sviluppi addirittura una forma di coscienza e relazionalità artificiale. Alla luce di queste riflessioni, possiamo dire che la distanza tra gli attuali LLM e AGI è ancora considerevole. Potrebbe ridursi col tempo, certo, ma per giungere a un vero salto qualitativo serve un cambiamento sostanziale, non semplicemente un aumento della complessità delle reti neurali. Di conseguenza, le paure di chi ritiene che la figura dello sviluppatore sia prossima all'estinzione rischiano di essere eccessive, basandosi sull'errata proiezione di uno sviluppo lineare e inarrestabile dell'AI verso l'ASI (Artificial Super Intelligence). Occorre, piuttosto, focalizzarsi su come integrare i progressi dell'intelligenza artificiale (debole, ma fortemente specializzata) all'interno delle pratiche di sviluppo del software, massimizzando i benefici e minimizzando i rischi. Già gli attuali LLM presentano questioni non banali di sicurezza e di affidabilità, le quali anticipano solo in parte i problemi ancor più delicati che sorgerebbero di fronte a un'AGI. Prima di preoccuparci di un ulteriore Godzilla ancora più grande, sarebbe opportuno imparare a gestire e a domare quello che abbiamo già davanti, definendo degli standard, linee guida e forme di vigilanza che garantiscano una convivenza costruttiva tra l'essere umano e l'AI.

9 Perché fermare l'AI non ha senso

L'innovazione tecnologica è un elemento inevitabile dello sviluppo umano. Dalla scoperta del fuoco alla rivoluzione digitale, ogni progresso tecnologico ha avuto un impatto significativo sulla società, portando sia benefici che problematiche. Tentare di fermare o rallentare l'innovazione tecnologica, in questo caso l'AI, è un'impresa inutile, semplicemente perché l'innovazione non può essere fermata, ogni innovazione ha sempre portato con sé più benefici che danni. La storia dell'umanità è piena di esempi di innovazioni che hanno rivoluzionato il modo in cui viviamo e lavoriamo. La rivoluzione industriale ha trasformato le economie e le società, e la rivoluzione informatica ha cambiato radicalmente il modo in cui comunichiamo, lavoriamo e interagiamo. Ogni nuova tecnologia ha creato nuove opportunità e nuove professioni, mentre alcune figure professionali tradizionali sono scomparse o si sono trasformate. Cercare di resistere all'innovazione tecnologica può avere conseguenze negative a lungo termine. Le aziende e le nazioni che tentano di bloccare l'adozione di nuove tecnologie rischiano di essere superate da concorrenti più lungimiranti che abbracciano il cambiamento e sfruttano le nuove opportunità offerte dalla tecnologia. Inoltre, la resistenza all'innovazione può portare a una perdita di posti di lavoro e a una diminuzione della produttività, poiché la maggior parte delle volte le tecnologie emergenti offrono soluzioni più efficienti e sostenibili. L'adozione dell'intelligenza artificiale offre numerosi benefici che vanno oltre la semplice automazione dei processi. L'AI può migliorare la qualità del lavoro e ridurre i costi operativi. Piuttosto che tentare di fermare l'AI, è essenziale sviluppare un quadro regolamentare che ne guidi l'uso in modo etico e responsabile. Dunque l'innovazione è un processo inarrestabile, e l'intelligenza artificiale (AI) ne è l'ultima manifestazione. Proprio come in passato abbiamo assistito a tentativi di bloccare tecnologie emergenti, come l'India che ha cercato di vietare la blockchain, è impossibile fermare il progresso tecnologico. Provare a vietare l'AI sarebbe come cercare di fermare il vento, possiamo fermare il vento? La risposta è semplice, non si può. Questo schema si ripete ciclicamente con ogni innovazione. Ogni 20-30 anni emerge una tecnologia esponenziale, inizialmente

ignorata, poi criticata, attaccata, e infine celebrata. E, come sempre, quando l'innovazione diventa indispensabile, tutti salgono sul carro dei vincitori, proclamando di aver sempre creduto in essa. Con l'AI, siamo già nel pieno di questo ciclo. L'iniziale scetticismo ha lasciato il posto a un uso diffuso e alla consapevolezza delle sue potenzialità, ma come è successo con internet e la blockchain ed altre tecnologie innovative, il percorso non è lineare. Internet ha impiegato circa 20 anni per essere compreso e adottato su larga scala, la blockchain ci ha messo circa 10 anni. Con l'AI, il processo sembra accelerare ulteriormente, evidenziando che più avanzano le tecnologie, più si riduce il tempo necessario per la loro diffusione e adozione. Tuttavia, questa velocità pone un rischio significativo. Stiamo parlando già di AI generativa, computer quantistici, economia spaziale, estrazione di risorse dagli asteroidi e altre innovazioni futuristiche, mentre molte persone non hanno ancora familiarità con strumenti basilari come la firma digitale o la posta elettronica certificata. Questo gap rischia di lasciare indietro intere fasce della popolazione, incapaci di tenere il passo con un'innovazione sempre più esponenziale. La sfida non è solo tecnologica, ma anche culturale ed educativa. Per evitare di ampliare questo gap, è necessario promuovere l'accessibilità e l'alfabetizzazione digitale, garantendo che tutti abbiano l'opportunità di salire sul carro dei vincitori di questa nuova tecnologia e dalle future tecnologie, senza essere lasciati indietro da un futuro che corre sempre più veloce.

10 Economia e Disruptive Marketing

È frequente imbattersi in interviste o dichiarazioni di CEO e figure di spicco che annunciano la fine della programmazione, enfatizzando l'imminente arrivo di robot e AI che potranno fare tutto meglio e più rapidamente dell'essere umano. Questi annunci hanno un forte impatto sull'opinione pubblica e sugli investitori generando un notevole ed immediato hype, queste previsioni sono spesso alimentate da un forte interesse economico. I colossi della tecnologia mirano a guadagnare credibilità per poi tradurre questa credibilità in valore azionario, investimenti e notorietà, dunque, in parole semplici, in soldi, pun-

tano a guadagnare sempre di più (beh, chi è che non lo farebbe). Nel caso di NVIDIA, la frase "Non studiate più programmazione" attribuita al CEO di NVIDIA Jensen Huang, ne è un esempio. Dichiarazioni estreme come questa, oltre a fare scalpore, hanno un impatto diretto sulla percezione del mercato e, di conseguenza, sul valore delle azioni. E infatti, non a caso, dopo queste affermazioni, il titolo NVIDIA ha registrato in maniera positiva un balzo significativo. Tuttavia, fare rumore fa parte di una strategia di marketing volta a stimolare l'interesse delle persone e degli investitori. Dunque, a livello pratico, l'idea di non studiare più programmazione è, ad oggi, priva di senso, questo perché la programmazione è e continuerà a essere la lingua con cui diamo istruzioni alle macchine. Potrà cambiare l'approccio, diventerà più astratto, ma saper programmare (magari a livelli sempre più alti di astrazione) resterà indispensabile. Pensiamo a un sistema AI che genera codice in modo autonomo. Chi controlla la correttezza del codice prodotto? Chi delinea gli obiettivi, chi definisce l'architettura, chi mette in sicurezza i dati? Tutte queste attività rientrano nel dominio della programmazione, intesa in senso più ampio rispetto al semplice scrivere righe di codice. Sostenere che tutto questo lavoro scompaia, o meglio, possa scomparire è, al momento, un puro slogan per generare hype, per far parlare di sé, e ci sono riusciti. Spesso, grandi aziende presentano prototipi spettacolari lasciando intendere che la rivoluzione tecnologica sia già una realtà compiuta. NVIDIA non è stato l'unico caso, ci sono tantissimi altri casi simili, alcuni casi più eclatanti sono Tesla Optimus 2, Amazon Go e ChatGPT o-1 e o-3:

- **Optimus 2 di Tesla:** Presentato come robot umanoide capace di muoversi autonomamente e compiere attività complesse, in realtà, molte funzioni che sembravano svolte dal robot erano comandate da persone fisicamente presenti o in remoto. Non si tratta di un fake totale, ma di un prototipo molto lontano dall'autonomia mostrata nella presentazione ufficiale.
- **Amazon Go e Amazon Fresh:** Nati come negozi senza casse, erano stati lanciati come esempi di completa automazione, ma si è scoperto

che varie persone svolgevano e svolgono tuttora controlli a distanza, verificando cosa i clienti avessero effettivamente acquistato. L'idea è rivoluzionaria, ma la tecnologia necessita ancora di anni di perfezionamento prima di funzionare senza supervisione umana.

- **ChatGPT o-1 e o-3 e l'illusione del pensiero:** L'introduzione delle versioni o-1 e o-3 di ChatGPT in grado di rispondere con frasi come "I'm thinking" prima di generare la risposta, ha generato l'idea di un'AGI già presente. In verità, stiamo osservando un raffinamento nel test time compute, dove invece di scegliere la parola più probabile in modo greedy (come i modelli precedenti), la generazione avviene valutando in parallelo diverse possibilità e confrontando vari scenari. Ricapitolando:

- **Modelli GPT di base:** La generazione avviene parola per parola, scegliendo ogni volta l'opzione con la probabilità più alta, in base a parametri come temperatura e numero di candidati.
- **Modelli o-1 e o-3:** Per ogni parola ipotizza più cammini di generazione, verifica quale porti a una frase finale più coerente e poi seleziona la migliore. Dunque, il Test Time Compute si traduce in un apparente ragionamento interno, ma non equivale alla capacità di pensare in senso umano.

C'è da sottolineare che non è tutto un bluff, le aziende stanno realmente investendo miliardi di dollari nell'AI, e i progressi ottenuti sono innegabili. Tuttavia, portare un prototipo a un livello di prodotto realmente autonomo e affidabile richiede anni, se non decenni. Nonostante ciò, il marketing delle grandi aziende sfrutta strategicamente il potere dell'hype, spesso ingigantendo il livello di autonomia e intelligenza effettiva delle tecnologie presentate. Siamo costantemente esposti a demo e presentazioni che generano innumerevoli investimenti e discussioni (positive e negative), creando aspettative che spesso si discostano dalla realtà immediata. Il disruptive marketing gioca un ruolo fondamentale in questo contesto, accelerando l'adozione di nuove tecnologie ma creando al tempo stesso illusioni. Ad oggi, molte aziende puntano su dichia-

razioni d'effetto per orientare le scelte di consumatori e investitori, spingendo verso l'idea che l'AI sia una soluzione magica a ogni problema. In realtà, il percorso verso una tecnologia ancora più avanzata richiede tempo, test approfonditi, correzione dei bias e un controllo etico rigoroso. Molte delle soluzioni presentate come completamente automatizzate dipendono ancora dal supporto umano per il loro funzionamento, vero che i risultati raggiunti sono notevoli, ma è cruciale mantenere uno sguardo critico. La dichiarazione riportata in precedenza come "Non studiate programmazione" possono colpire molte persone, ma non rispecchiano il panorama reale, il mercato del software e il ruolo dello sviluppatore stanno crescendo in complessità e importanza, espandendosi verso nuovi campi di applicazione e non scomparendo. Oggi giorno dobbiamo anche saper distinguere tra i progressi reali e le promesse ancora in fase sperimentale, senza sminuire l'innovazione, ma evitando di sopravvalutarne i tempi di realizzazione.

11 Il futuro dello sviluppatore: Creatore non solo orchestratore

Formulare previsioni, sia nel breve che nel lungo termine, è un'operazione estremamente complessa. Anche cercare di immaginare cosa accadrà tra un secondo o un minuto può rivelarsi difficile, e questa complessità aumenta esponenzialmente quando si parla di scenari a 10, 50 o 100 anni nel campo tecnologico. Tuttavia, immaginare il futuro non è solo un esercizio intellettuale, ma un'attività essenziale. Anche le previsioni più incerte aiutano a stabilire priorità, orientando la ricerca e gli investimenti verso settori chiave per il progresso della società. Inoltre, riflettere sui possibili scenari consente di anticipare questioni etiche, sociali ed economiche, favorendo una preparazione più consapevole per affrontare le sfide future. In questo contesto, proverò a fornire una mia interpretazione dell'impatto degli LLM e del futuro degli sviluppatori nei prossimi 10, 50 e 100 anni.

11.1 Tra 10 anni: sviluppatore progettista di strumenti AI

Nel prossimo decennio, gli sviluppatori non si limiteranno più soltanto a scrivere codice, il ruolo dello sviluppatore assumerà un carattere sempre più orientato alla progettazione di strumenti che facilitino la collaborazione tra esseri umani e sistemi AI. Se oggi gli sviluppatori si occupano principalmente di scrivere codice e integrare servizi, nei prossimi dieci anni diventeranno figure strategiche capaci di:

- **Creare interfacce di interazione naturale:** Gli sviluppatori dovranno progettare non solo la logica di business del software, ma anche l'esperienza utente (UX) in un modo ancora più evoluto. Bisognerà rendere il dialogo trasparente con gli LLM e gli altri sistemi AI. Non si tratterà più solo di widget, di pagine web o di un'applicazione, bensì di ambienti virtuali dove l'utente potrà dialogare.
- **Gestire etica e protezione dei dati:** Già nei prossimi dieci anni, sarà inevitabile confrontarsi con normative sempre più restrittive nel contesto AI. Lo sviluppatore sarà incaricato di tradurre i principi etici e le leggi in vincoli concreti all'interno dei sistemi AI, garantendo il rispetto dei diritti individuali, la trasparenza delle decisioni e l'assenza di discriminazioni. Sarà più che mai necessario garantire un uso rispettoso dei dati, considerando i rischi legati alla privacy e alle possibili manipolazioni dei modelli. In questa prospettiva, lo sviluppatore assumerà un ruolo di custode di come le informazioni vengono raccolte, immagazzinate e utilizzate.
- **Integrare processi di validazione e controllo della qualità:** Creare e utilizzare LLM comporta definire metriche di qualità, ossia, coerenza, correttezza e pertinenza. Gli sviluppatori controlleranno la qualità dei contenuti generati dall'AI, definendo sistemi di validazione automatica e parametri di misurazione. Più l'AI diventa autonoma, più sarà importante garantire controlli e verifiche continue. Questo comporterà l'uso di

algoritmi per l'analisi automatica di errori o di bias. Dunque, mentre una parte del codice sarà generata automaticamente dall'AI, lo sviluppatore dovrà esercitare la capacità di verificare, correggere e adattare le soluzioni proposte. Paradossalmente, l'AI renderà necessario un maggiore numero di figure competenti per integrare e testare i sistemi e garantire la sicurezza.

- **Ottimizzare i prompt:** L'arte di formulare richieste (il cosiddetto prompt engineering) diventerà fondamentale. Gli sviluppatori non solo scriveranno codice, ma svilupperanno anche strumenti che permettano di insegnare all'AI come interpretare domande, contesti e finalità, ottimizzandone le risposte e riducendo il rischio di generazioni inadeguate.

In breve, a 10 anni da oggi, la professione dello sviluppatore si concentrerà su strumenti e processi che semplifichino l'uso dell'AI da parte di qualunque utente. Vedremo la nascita di un nuovo tipo di progettista di strumenti AI, che creerà collegamenti tra le capacità dell'intelligenza artificiale e le necessità pratiche, creative ed etiche dell'essere umano.

11.2 Tra 50 anni: sviluppatore come creatore di soluzioni per un mondo inclusivo e sostenibile

Allungando l'orizzonte temporale a cinquant'anni, potremo trovarci in un contesto in cui l'AI sarà pervasiva in ogni settore della società. In questo scenario, lo sviluppatore non svolgerà soltanto il ruolo di progettista di strumenti AI, ma diventerà un vero e proprio innovatore per la risoluzione di problemi globali, esso dovrà:

- **Affrontare sfide planetarie:** Il cambiamento climatico, la carenza di risorse, l'assistenza sanitaria universale saranno alcune tra le sfide centrali. Gli sviluppatori del futuro si troveranno a collaborare con reti di AI specializzate, capaci di processare quantità enormi di dati per elaborare soluzioni. L'infrastruttura hardware e computazionale di modelli AI enormi incide sul pianeta, basti pensare alle spese energetiche dei grandi

data center. Lo sviluppatore del futuro dovrà progettare sistemi efficienti, riducendo consumi, riciclando componenti e minimizzando l'impronta ecologica. Saranno necessarie competenze di ottimizzazione, oltre a una visione del problema dall'alto che includa valori di responsabilità ambientale.

- **Creare ecosistemi di AI multidisciplinari:** Entro cinquant'anni, non avremo una singola super AI, ma una costellazione di AI specializzate in settori come medicina, climatologia, giuridico, tutte progettate per collaborare tra loro. La figura dello sviluppatore non sarà quella di un semplice orchestratore che si limita a coordinare sistemi preesistenti, ma di un autentico creatore, capace di immaginare e costruire ecosistemi integrati. Il suo lavoro andrà oltre il semplice collegamento di strumenti diversi, sarà incentrato sulla creazione di interazioni tra le intelligenze artificiali, favorendo una cooperazione tra discipline che permetterà di risolvere problemi ancora più complessi e di generare soluzioni innovative che altrimenti sarebbero irraggiungibili senza l'interazione di più AI insieme.
- **Promuovere inclusività e accessibilità:** In cinquant'anni, avremo AI integrate in ogni aspetto della vita quotidiana. Toccherà agli sviluppatori progettare sistemi inclusivi, capaci di interagire con persone che presentano diverse abilità fisiche, sensoriali o cognitive. L'attenzione all'accessibilità non sarà più un capitolo facoltativo, ma un requisito fondamentale per qualsiasi progetto tecnologico, creando soluzioni che possano migliorare la vita di milioni di persone, riducendo disuguaglianze e disagi.
- **Innovare i metodi di sviluppo:** Oggi parliamo di DevOps come integrazione tra sviluppo e operazioni. Fra cinquant'anni, potremmo invece parlare di un AI-DevOps evoluto, dove l'AI stessa parteciperà alla scrittura, al testing e al deployment del codice. Lo sviluppatore diventerà più un direttore d'orchestra della produzione software, con compiti di supervisione e raffinamento, sebbene il termine orchestratore finirà gradual-

mente per lasciare spazio a quello di creatore, poiché la responsabilità e la visione di insieme resteranno caratteristiche e qualità umane.

In breve, tra 50 anni, lo sviluppatore sarà il creatore di ecosistemi complessi e sostenibili, capaci di integrare AI specializzate per affrontare problemi ancora più complessi. Non si limiterà a progettare tecnologie, ma guiderà innovazioni che promuovano inclusività, sostenibilità e cooperazione multidisciplinare.

11.3 Tra 100 anni: sviluppatore come custode dell'umanità digitale

Spingendoci a un secolo di distanza, entriamo nel regno della speculazione, possiamo immaginare scenari oggi difficili anche solo da concepire. Tuttavia, è utile tentare uno sguardo di lungo termine. In tale mondo, lo sviluppatore assumerebbe un ruolo quasi filosofico:

- **Custode dell'etica digitale:** I valori e i principi umani dovranno essere preservati in una realtà in cui AI e integrazione tecnologica saranno la norma. La figura dello sviluppatore si tramuterà in quella di un custode dell'etica, incaricato di garantire che le azioni e le decisioni delle AI rispettino la dignità umana e l'ecosistema in cui viviamo.
- **Governance interplanetaria:** Stando alle previsioni di Elon Musk, CEO di SpaceX, è plausibile che le questioni etiche e decisionali non riguarderanno più solo l'ambito terrestre. In un futuro a cent'anni, l'umanità potrebbe avere colonie spaziali o basi su altri pianeti. In tal caso, i sistemi AI gestiranno risorse a enormi distanze, e gli sviluppatori dovranno definirne le regole di funzionamento e di interazione con gli umani, in una prospettiva che va ben oltre la Terra.
- **Fusione tra uomo e macchina:** Con la diffusione delle interfacce cervello-computer, parti fondamentali della conoscenza e dell'identità umana come ricordi, saperi, coscienza e pensiero, potrebbero essere trasferite o condivise direttamente con le macchine. In questo contesto, gli sviluppatori non saranno più semplicemente creatori di software, ma veri

e propri specialisti nell'integrazione tra uomo e macchina, assumendosi la responsabilità di proteggere i confini tra ciò che è personale e ciò che è digitale. Questo significherà anche garantire che tali tecnologie rispettino i valori e i diritti fondamentali dell'essere umano, questo perché man mano che la tecnologia diventerà sempre più invasiva, la fragilità umana potrebbe trovarsi di fronte ad una fusione genetica tra uomo e macchina, fino alla possibilità di controllare i pensieri stessi. Questa ipotesi potrebbe sollevare dei dubbi sulla natura dell'essere umano, che potrebbe non essere più puro come lo intendiamo oggi. La fusione tra uomo e macchina potrebbe trasformare radicalmente l'identità umana, con alcuni che potrebbero scegliere di mantenere la loro natura biologica, mentre altri potrebbero optare per un'esistenza ibrida o completamente digitale, mettendo in discussione il concetto stesso di umanità.

In un mondo così profondamente trasformato, la figura dello sviluppatore assumerà un ruolo filosofico e custode dei valori umani. Dovrà interrogarsi su cosa significhi davvero essere umano in un'era in cui i confini tra il biologico e il tecnologico si fanno sempre più sfumati. Sarà inevitabile stabilire non solo come, ma anche se queste tecnologie debbano essere integrate nella società. In questo scenario, lo sviluppatore non sarà solo un tecnico, ma un custode dei valori e dei significati che definiscono l'umanità stessa.

11.4 È più facile prevedere ciò che (forse) non accadrà

Se prevedere eventi positivi con certezza può essere complicato, spesso risulta più semplice escludere determinate possibilità applicando un ragionamento per assurdo. Come in matematica, per dimostrare una verità si utilizza spesso la dimostrazione per assurdo, si parte dall'ipotesi opposta a quella che si vuole dimostrare, mostrando che conduce a una contraddizione. Questo metodo consente di arrivare a conclusioni in modo più semplice. Allo stesso modo, nella realtà, possiamo applicare un ragionamento simile, invece di concentrarci su ciò che può accadere, possiamo escludere ciò che non può accadere:

- **Non avremo un'AI onnisciente che capirà tutto e tutti:** Anche immaginando l'esistenza di sensori e modelli ultra avanzati, il concetto di tutto rimane estremamente difficile, se non impossibile, da raggiungere. Neppure noi esseri umani, che per alcuni filosofi siamo creature perfette, siamo in grado di comprendere tutto e tutti. La nostra comprensione si basa su esperienze soggettive e su una coscienza che le macchine, per quanto sofisticate, non possiedono né possiamo sapere se mai riusciranno a sviluppare.
- **Non avremo un mondo senza sviluppatori oppure professionisti IT:** Pensare che la tecnologia si autogestisca al punto di eliminare la necessità di competenze umane non è assolutamente uno scenario plausibile. C'è sempre bisogno di chi definisce gli obiettivi, chi controlla gli impatti e chi adegua costantemente i sistemi a nuovi contesti.
- **Non scomparirà la creatività e l'intuizione umana:** Sebbene i modelli generativi possano diventare estremamente abili nel proporre idee, resta difficile immaginare che possano sostituire l'atto creativo puro ed intuitiva degli esseri umani.

12 Conclusioni

Eccoci qui, siamo arrivati verso la fine di questa relazione, dove abbiamo analizzato in modo sistematico la natura del pensiero umano, confrontandolo con le capacità degli LLM. Abbiamo visto come il pensiero umano sia unico, caratterizzato da coscienza, intenzionalità ed esperienza vissuta, elementi non replicabili dall'AI, la quale opera su basi probabilistiche e manca di reale comprensione e di come gli LLM falliscano nel generare vera innovazione, limitandosi a rimescolare pattern esistenti. Abbiamo affrontato i rischi principali legati agli LLM, come i bias, le complessità etiche e i limiti di una possibile etica programmabile. Abbiamo considerato le opportunità offerte dall'AI, come la Neuro-Symbolic AI e i modelli privi di dati, e le sfide dell'AGI, sottolineando che fermare l'AI sarebbe inutile e dannoso. Infine ci siamo proiettati nel

futuro, prevedendo come il ruolo dello sviluppatore sia destinato a evolvere, da progettista di strumenti AI nei prossimi 10 anni, a creatore di soluzioni per un mondo più inclusivo e sostenibile tra 50 anni, fino a diventare, tra 100 anni, il custode dell'umanità digitale, capace di preservare i valori umani in un mondo sempre più tecnologico.

12.1 L'esperimento finale

Come anticipato all'inizio, questa relazione include un piccolo esperimento destinato ai lettori di questa relazione. La domanda che ti pongo è: "Saresti più sorpreso se questa relazione fosse stata scritta da un'AI oppure da un essere umano?" E, nel caso avessi risposto "da un'AI", ti chiedo: "Che percentuale di questo testo credi sia stata generata da un LLM?"

La risposta, probabilmente, ti sorprenderà, 0%. Esatto, l'intero contenuto di questa relazione è stato scritto senza alcun intervento da parte di un modello di intelligenza artificiale. Te lo saresti aspettato? La motivazione che si cela sul perché ho scelto di non utilizzare l'AI è per una ragione ben precisa, durante l'ultima lezione sugli LLM, ho provato un senso di incertezza, quasi di smarrimento, davanti all'idea che l'intelligenza artificiale potesse un giorno rendere obsoleto il ruolo degli sviluppatori. Mi sono chiesto quale sarebbe stato il valore del mio percorso di studi in un mondo in cui l'AI sembra capace di fare tutto. Questa relazione è stata una sfida per riconquistare fiducia nelle mie capacità. Scriverlo autonomamente è stato un modo per dimostrare a me stesso che, nonostante i progressi dell'AI, la creatività, l'intuizione e la riflessione critica proprie dell'essere umano rimangono insostituibili. Più di un semplice progetto accademico, questo è stato un atto di riaffermazione del valore delle competenze umane, un modo per dimostrare che, anche di fronte a tecnologie avanzate, la capacità di creare valore autentico risiede ancora nell'uomo. Durante la stesura, ho raccolto informazioni e ispirazioni da fonti come video YouTube, post su Reddit, post su Instagram, articoli sull'AI, pagine Wikipedia e ricerche online. Questo processo mi ha permesso di riorganizzare le idee, reinterpretarle e integrarle in un discorso coerente, dimostrando che l'apporto

umano va ben oltre il semplice assemblaggio di informazioni.

12.1.1 Poniamo la stessa domanda a ChatGPT

La relazione nasce da una domanda fondamentale: "L'intelligenza artificiale eliminerà il lavoro del programmatore?". Propongo ora di rivolgere la stessa domanda a ChatGPT. Di seguito, mostro il prompt utilizzato e la risposta generata in tempo reale dal modello:

- **Prompt:** "L'Intelligenza Artificiale eliminerà il lavoro del programmatore?"
- **Risposta ChatGPT:** "La risposta breve è: no, ma cambierà il modo di lavorare dei programmatori.

L'impatto dell'Intelligenza Artificiale (IA) sul mondo del lavoro, e in particolare su quello dei programmatori, è una domanda che suscita grande interesse e dibattito. È vero che l'IA sta avanzando a passi da gigante e sta già automatizzando molte attività che prima richiedevano l'intervento umano. Tuttavia, affermare che l'IA eliminerà completamente il lavoro del programmatore è un'affermazione troppo semplicistica e probabilmente errata.

Ecco perché:

- **L'IA come strumento, non come sostituto:** L'IA è destinata a diventare uno strumento potentissimo nelle mani dei programmatori, piuttosto che un loro sostituto. Molte attività ripetitive e routinarie potranno essere automatizzate, liberando i programmatori per concentrarsi su compiti più creativi e complessi, come la progettazione di architetture software innovative, la risoluzione di problemi complessi e lo sviluppo di algoritmi avanzati.
- **La necessità di competenze umane:** Anche se l'IA può generare codice, è ancora necessaria l'intelligenza umana per valutare la qualità del codice generato, per identificare eventuali errori e per prendere decisioni strategiche sullo sviluppo del software. Inoltre,

l'IA ha difficoltà a comprendere i contesti complessi e a gestire situazioni impreviste, che richiedono un approccio flessibile e creativo tipico dell'essere umano.

- **Evoluzione dei ruoli:** Piuttosto che eliminare i lavori, l'IA potrebbe portare a una ridefinizione dei ruoli dei programmatori. I programmatori del futuro dovranno possedere competenze sempre più avanzate nell'ambito dell'IA, del machine learning e della data science, per poter interagire efficacemente con questi strumenti e sfruttarne al meglio le potenzialità.

In conclusione, mentre è indubbio che l'IA avrà un impatto significativo sul mondo della programmazione, è improbabile che elimini completamente questa professione. Piuttosto, l'IA trasformerà il modo in cui i programmatori lavorano, rendendo il loro lavoro più efficiente e creativo. Per i programmatori, questo significa la necessità di acquisire nuove competenze e di adattarsi a un mondo del lavoro in continua evoluzione.

Confrontando le due argomentazioni, ossia, la mia relazione e la risposta di ChatGPT, potrai osservare alcune differenze fondamentali. Non serve confrontare la velocità della generazione, un LLM sarà certamente estremamente rapido rispetto a un essere umano nel produrre una risposta strutturata e coerente. Tuttavia, ciò che diventa evidente è la mancanza di profondità e ampiezza del ragionamento. Un LLM si limita a rielaborare e riformulare informazioni già presenti nei suoi dati di addestramento, senza la capacità di sviluppare nuove idee o di esplorare a fondo i contesti complessi che una domanda di questo tipo implica. La sua visione puramente statistica, gli impedisce di offrire intuizioni originali o uno stile personale. Mentre può apparire linguisticamente impeccabile, manca della creatività. Invito voi lettori a fare lo stesso esperimento personalmente, ponete la domanda "L'Intelligenza Artificiale eliminerà il lavoro del programmatore?" a un LLM di vostra scelta. Analizzate la risposta e confrontatela con questa relazione. Concentratevi su aspetti come la capacità di approfondire il problema, la coerenza nel collegare concetti complessi e l'eventuale presenza di spunti nuovi o creativi. Voglio specificare

che questo confronto non serve a decretare un vincitore tra uomo e macchina, ma a riflettere su come i due approcci, umano e LLM possano completarsi a vicenda. L'obiettivo è comprendere quanto il contributo umano, fatto di creatività, riflessione critica e intenzionalità, resti ancora oggi insostituibile, anche di fronte a strumenti tecnologici sempre più potenti. Naturalmente, i modelli AI sono di grandissima utilità e possono velocizzare la scrittura o la ricerca di informazioni, ma affidarsi a loro in modo esclusivo rischia di appiattire il risultato, rinunciando alla sfumatura e al punto di vista che solo umano possiede.

12.2 Considerazioni finali

L'avvento degli LLM e delle tecnologie di intelligenza artificiale non rappresentano una minaccia per la figura dello sviluppatore, ma piuttosto una trasformazione che potenzia le sue capacità e amplia il suo ruolo nel processo di sviluppo software. Invece di temere l'AI, dovremmo abbracciarla come un'opportunità per ampliare le nostre competenze. Agli inizi di questa relazione, avevo mostrato un'immagine presentata durante l'ultima lezione sugli LLM, ovvero, di Godzilla che distruggeva le torri che simboleggiava la presunta fine del lavoro dello sviluppatore causata dagli LLM. Voglio concludere la relazione mostrando un'immagine della mia interpretazione, presentando invece la versione opposta, ossia, un Godzilla che non distrugge più, ma collabora. L'AI ci offre forza e potenza, noi offriamo umanità, etica e visione. In questa stretta di mano, nessuno dei due scompare, entrambi diventiamo indispensabili per costruire il futuro. In questa immagine gli LLM sono rappresentati sempre da Godzilla poiché, se impiegata male, un'AI potente può provocare danni enormi, l'idea è mostrare che la tecnologia, se usata con visione e responsabilità, diventa un partner anziché un nemico. Non c'è ostilità, ma sinergia tra Godzilla (LLM) e l'uomo. Inoltre, le torri intatte simboleggiano le competenze e le professioni che, ben lontane dall'esser distrutte, vengono supportate e amplificate dalla potenza dell'AI.

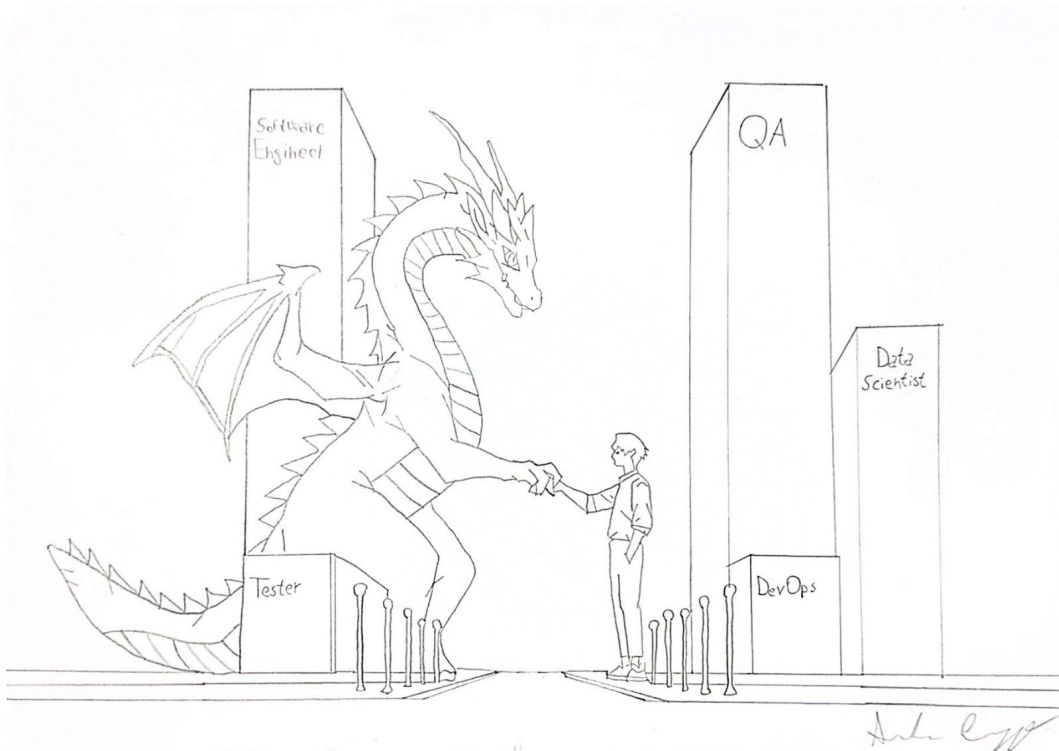


Figura 8: La mia personale interpretazione

Infine, la narrazione catastrofista che vedeva gli LLM come un Godzilla inarrestabile pronto a demolire gli sviluppatori ignorava il fatto che, nella realtà, la tecnologia è un moltiplicatore di potenzialità, non un sostituto integrale dell'essere umano. Il senso di responsabilità e la capacità di introspezione sono tratti umani non sostituibili, pertanto lo sviluppatore sta cambiando, non scomparendo. Lo sviluppatore è immortale.