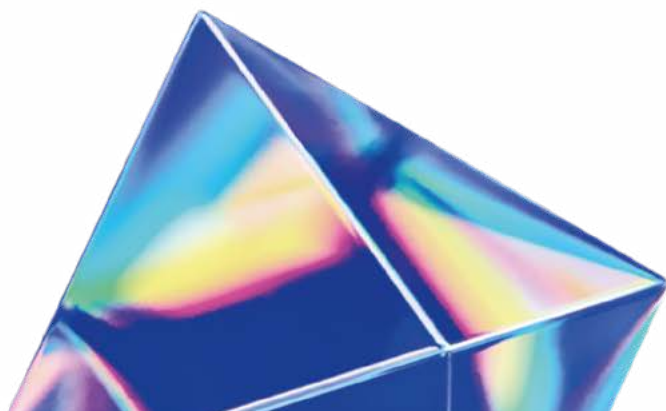
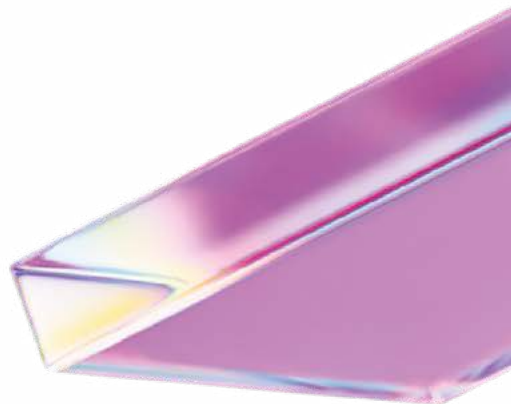
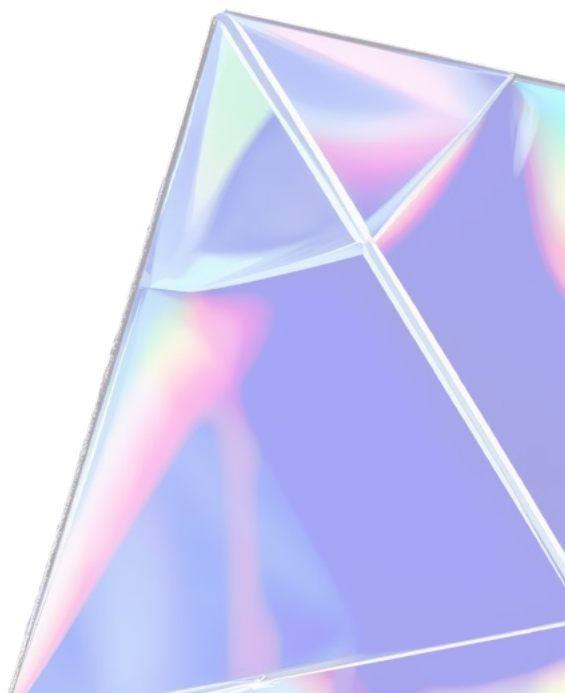
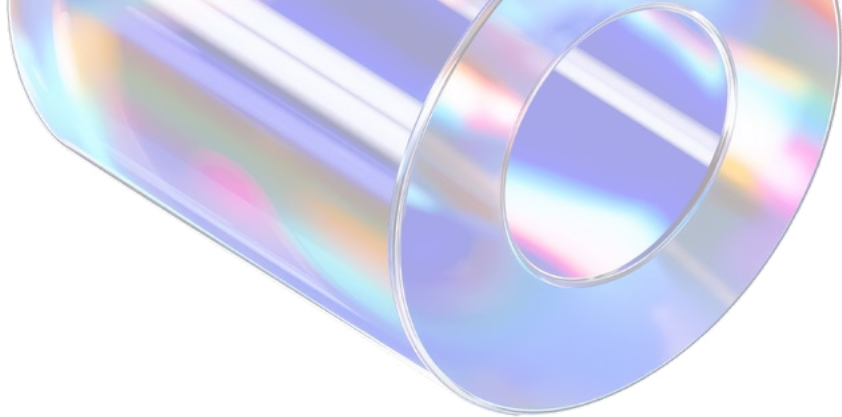




L'Intelligenza Artificiale Generativa
per l'impresa: una guida pratica

G.A.I.A.





La trasformazione digitale e tecnologica rappresenta un elemento prioritario per lo sviluppo dell'economia e delle imprese.

La digitalizzazione può, infatti, aumentare la competitività in tutti i settori produttivi, sostenendo la crescita del tessuto economico ed uno sviluppo più sostenibile.

Per tutte le imprese, ed in particolare per le PMI, è un fattore sempre più importante e strategico su cui puntare. Il tema della competitività è cruciale per le nostre PMI; gli strumenti digitali consentono di ottimizzare i processi, realizzare nuovi prodotti e servizi ed espandersi in nuovi mercati, sia in una logica di export che di vera e propria internazionalizzazione.

Trasformazione digitale, governo della sicurezza informatica ed Intelligenza Artificiale sono aspetti essenziali e sinergici per una maggiore efficienza, produttività e presenza efficace sui mercati globali.

Il cammino verso una piena maturità nella trasformazione digitale delle PMI italiane richiede una stretta collaborazione tra le imprese e chi, come **MISTER Smart Innovation**, si occupa di ricerca industriale, accelerazione e trasferimento tecnologici.

MISTER è un laboratorio di ricerca industriale e centro per l'innovazione accreditato alla Rete Alta Tecnologia dell'Emilia-Romagna.

Le attività principali di MISTER includono azioni di ricerca industriale, sviluppo sperimentale, prototipazione, trasferimento tecnologico, divulgazione scientifica e Open Innovation. Inoltre, MISTER è soggetto gestore del Tecnopolo Bologna CNR. La società opera in ambiti tecnologici avanzati quali sensoristica industriale e ambientale, Intelligenza Artificiale e machine learning, realtà aumentata e simulazioni virtuali, additive manufacturing e progettazione 3D e nanotecnologie e biosensori.

MISTER, in qualità di sportello della Rete Regionale per la Transizione Digitale delle imprese, e attraverso il progetto **GAIA - Generative AI per Imprese e Assessment**, mette a disposizione strumenti, servizi e contenuti per le imprese.

La pubblicazione, "**L'Intelligenza Artificiale Generativa per l'impresa: una guida pratica**" rappresenta un'iniziativa volta a rendere chiara e accessibile l'evoluzione dell'Intelligenza Artificiale, evidenziandone gli impatti sulle attività produttive e offrendo una prospettiva sui futuri sviluppi.

Il direttore

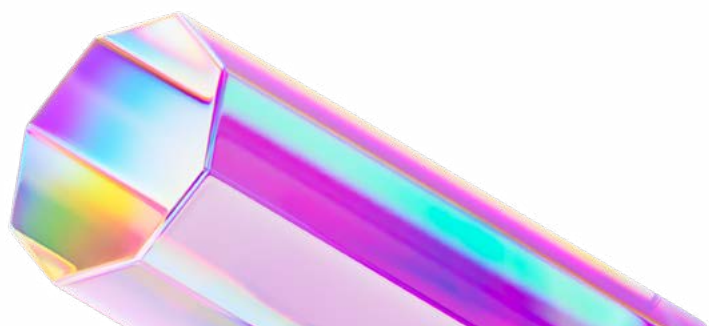
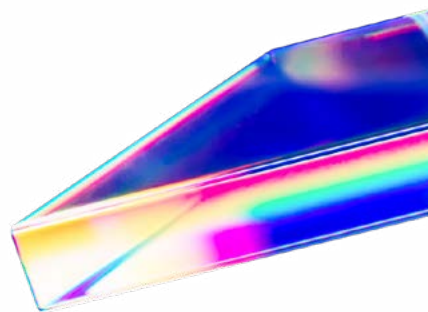
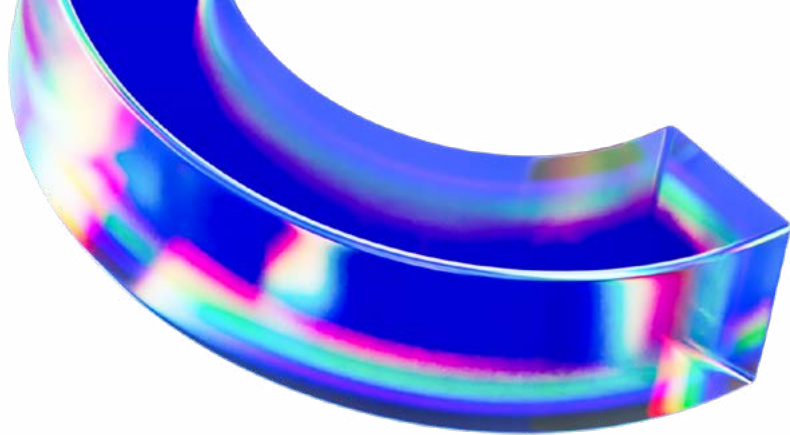
Valeria Pignedoli

「Indice」

1. Uno strumento di uso quotidiano	6
IA in azienda	9
Qualcosa da migliorare	10
2. Usare gli LLM in azienda	12
L'IA "consulente"	15
L'IA "tirocinante"	18
L'IA web software	22
L'IA "ricercatore"	24
L'IA "agente"	27
Tutti i pesci del mare	28
Qualche spunto ulteriore	30
Multimodalità	30
Documenti	30
L'arte del prompting	31
Attenti alla privacy!	33
3. L'IA generativa come sistema aziendale	34
Identikit dello strumento IA di successo	38
Parole chiave per l'IA generativa fatta sistema	40
4. Conclusioni	48

01

「Uno
strumento
di uso
quotidiano」



L'Intelligenza Artificiale (IA) si è diffusa con una rapidità senza precedenti ed è ormai di uso quotidiano, segnando un punto di svolta per l'intero mondo digitale. Sebbene siano diversi i campi dell'IA soggetti a questa evoluzione repentina, l'aspetto di maggiore impatto sulla cultura di massa, affermandosi nella vita di tutti i giorni di milioni di persone, è quello legato all'**Intelligenza Artificiale Generativa (GenAI)**.

Con il termine GenAI si indicano i sistemi capaci non solo di analizzare dati esistenti, ma di creare contenuti (testi, codice, immagini, suoni) che imitano la creatività umana. Strumenti basati su **Large Language Model (LLM)** come ChatGPT, Gemini o Claude, sono gli esempi più noti di questa tecnologia e comunemente usati dagli utenti.

A seguito della loro enorme popolarità, questi sistemi vengono ora rapidamente adottati anche nel mondo del lavoro. Le imprese, dalle startup alle grandi industrie, vedono nella GenAI la promessa di una maggior efficienza, di una maggiore capacità innovativa e di un decisivo vantaggio competitivo.



Glossario

IA: Intelligenza Artificiale, capacità di un sistema informatico di simulare aspetti dell'intelligenza umana quali il ragionamento, la comprensione e la creatività.

LLM: modello linguistico di grandi dimensioni, per la comprensione e generazione di testi. Esempi più noti sono GPT, Gemini e Claude.

Chatbot: interfaccia utente comunemente usata per l'interazione con gli LLM più diffusi. Simula una conversazione in chat con il sistema di IA.

IA in azienda

Per comprendere la portata del fenomeno, è sufficiente guardare ai dati. Uno studio della Wharton University e di GBK Collective¹, che definisce il 2025 come “l’anno dell’adozione” dell’IA su larga scala, riporta che l’uso settimanale di strumenti di GenAI da parte dei **decision maker aziendali** è passato dal 37% nel 2023 al 72% del 2025. Secondo un altro studio, l’*IA Index Report 2025* di Stanford², il tasso di adozione complessivo dell’IA nelle aziende ha visto una crescita del **23%** nel solo 2024, con la GenAI come motore primario di questa accelerazione. A livello di **investimenti** su scala globale, i privati hanno speso **33,9 miliardi** di dollari nel 2024 in tecnologie GenAI, segnando un aumento del 18,7% rispetto all’anno precedente.

Questo fermento si riflette anche in contesti locali particolarmente dinamici. Ne è un esempio **l’Emilia-Romagna**, che grazie ad una strategia di lungo termine fondata sull’ecosistema dell’innovazione e alla convergenza strategica della rete dei Tecnopoli della Regione Emilia-Romagna (di cui fanno parte infrastrutture di supercalcolo di livello mondiale) ha reso il territorio fertile per la transizione digitale. Qui, il mercato dell’IA ha raggiunto i **48,5 milioni** di euro nel 2023 e si prevede che **triplicherà** entro il 2027³. Il tasso di crescita regionale supera significativamente le medie nazionali e l’adozione da parte delle imprese sta accelerando, segnalando un cambiamento nelle strategie di investimento e innovazione.

¹ <https://www.gbkcollective.com/thoughtleadership/growing-up-navigating-gen-ais-earlyyears>

² <https://hai.stanford.edu/ai-index/2025-ai-index-report?utm>

³ <https://www.confind.emr.it/news/intelligenza-artificiale-emilia-romagna-mercato-485-milioni-di-euro-nel-2023>

Qualcosa da migliorare

Nonostante il grande fermento attorno all'IA generativa e gli investimenti senza precedenti che stiamo vedendo a livello globale, la realtà operativa è spesso più complessa di quanto sembri.

Da un lato molte aziende stanno già sperimentando e integrando la GenAI nei propri processi produttivi, e stanno già beneficiando di cambiamenti non solo in termini qualitativi ma anche in termini quantitativi. La vera sfida è mettere a sistema questi nuovi processi. Nel report *The State of AI in Business 2025* del MIT¹, questo scollamento viene definito “GenAI divide”: secondo lo studio, il 95% dei progetti pilota GenAI fatica a raggiungere gli obiettivi economici fissati.

Nonostante ciò, emergono segnali incoraggianti. I settori **tecnologico e media e telecomunicazioni** stanno già mostrando **cambiamenti strutturali** significativi grazie alla GenAI, mentre altri – servizi finanziari, sanità, manifattura – sono ancora in una fase più **esplorativa**. È un panorama eterogeneo ma capace di esprimere un chiaro potenziale, mostrando come con un corretto accompagnamento dell'innovazione si giunge a risultati concreti.

Anche **in Emilia-Romagna** assistiamo a una tendenza simile. Le **grandi aziende** del territorio, spesso leader nei rispettivi settori, si stanno distinguendo come esempi virtuosi, integrando soluzioni **IA specializzate** nei propri processi e dimostrando che la trasformazione è possibile e già in corso.

Le **Piccole e Medie Imprese**, invece, si trovano ad affrontare maggiori difficoltà. Mancano **punti di riferimento** chiari per orientarsi tra le molte soluzioni disponibili e, allo stesso tempo, le competenze richieste **evolvono** rapidamente. Anche gli **investimenti** necessari possono apparire eccessivi. È una situazione comprensibile, soprattutto in un momento in cui il mercato cambia a una **velocità senza precedenti**.

¹ https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf

Superare questi limiti, soprattutto per le PMI, richiede non solo capire **quali tecnologie** adottare, ma anche **quali approcci** evitare. I dati mostrano una dinamica interessante: pochi progetti IA personalizzati arrivano a una piena produzione, mentre dall'altra parte il **90% dei lavoratori** usa già regolarmente LLM tramite account personali. Strumenti generici come ChatGPT si stanno ritagliando uno spazio crescente anche nelle aziende, al punto che il 40% ha già sottoscritto **abbonamenti** formali.

Un motivo è chiaro: i sistemi aziendali personalizzati spesso non integrano memoria contestuale e quindi **non evolvono nel tempo**, mentre gli LLM di uso comune offrono **esperienze più fluide** e immediatamente utili. È naturale, quindi, che molti lavoratori e molte imprese scelgano di affidarsi a soluzioni già mature, stabili e aggiornate di continuo.

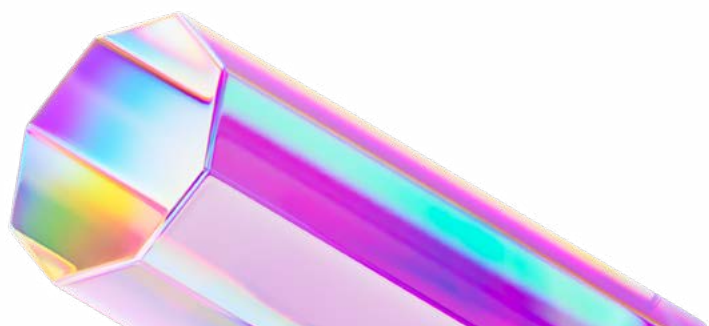
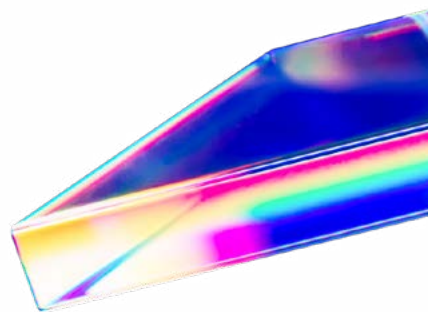
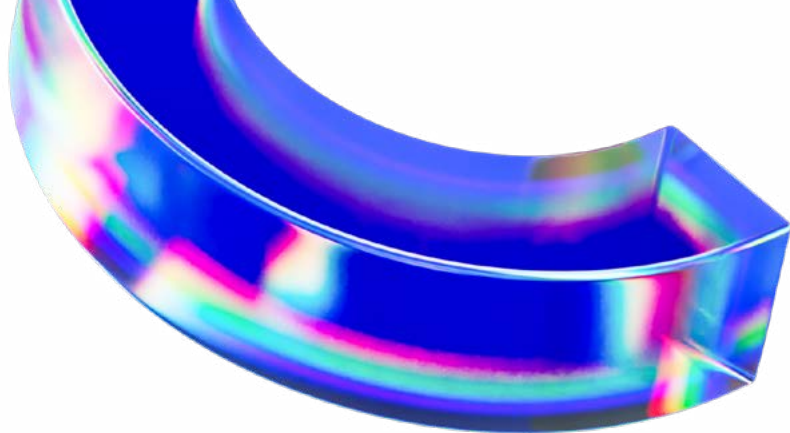
Se i sistemi su misura procedono lentamente e gli LLM di largo consumo prosperano, l'approccio più pragmatico ed efficace è imparare a usare bene proprio questi strumenti.

Da questa consapevolezza nasce la presente guida: un vademecum pensato per aiutare chiunque a sfruttare gli LLM nel modo corretto, anche senza competenze tecniche, trasformando la loro adozione in un **vantaggio competitivo** reale.

In queste pagine troverete strumenti pratici per muovere i primi passi con l'IA generativa, indicazioni sui sistemi più diffusi e consigli per usarli in modo efficace, evitando gli errori più comuni. Mostreremo anche quali percorsi può intraprendere un'azienda che voglia integrare l'IA in modo più stabile e strategico nelle proprie attività.

02

「Usare
gli LLM
in azienda」



I chatbot IA basati su LLM, di cui **ChatGPT** è l'esempio più noto, rappresentano oggi il **metodo più popolare** per utilizzare l'IA generativa per l'utente comune. Accessibili, facili da usare e versatili, possono generare moltissimi contenuti in meno tempo di quanto non siamo capaci di leggerli, offrendo immense possibilità.

Rimangono però strumenti controversi, circondati da una certa cattiva fama, non del tutto immeritata. I contenuti generati dall'IA sono ben scritti, piacevoli a primo impatto ma non hanno nessuna garanzia di correttezza e affidabilità. L'IA è utile, quindi, non se ci fidiamo ciecamente di ciò che scrive ma piuttosto se impariamo a conoscerla e a utilizzarla per quella che è: un software molto potente e non un genio digitale.

Di questi software ne esistono diversi. I modelli di **OpenAI** sono lo standard di fatto e **ChatGPT** è uno dei servizi più versatili e capaci. **Google** si è nel tempo guadagnato un rispetto analogo per i propri modelli della famiglia **Gemini**. **Claude** è un'alternativa di prestigio, sviluppata con particolare attenzione ai professionisti e alla produttività. Non tutti i modelli di IA sono uguali. OpenAI (ChatGPT), Google (Gemini), o Anthropic (Claude) hanno sviluppato **modelli IA diversi**, basandosi su tecnologie e dati di addestramento differenti. Ogni modello ha le proprie particolarità, e potrebbe dare **risposte diverse alla stessa domanda**: uno può eccellere nella scrittura creativa, un altro nell'analisi di codice o dati e, un altro ancora, nel fornire risposte più aggiornate o fattuali. Si tratta peraltro di differenze spesso sottili: ad esempio Claude riesce a scrivere codice funzionante e di qualità, ma non necessariamente più completo di quello prodotto da ChatGPT o Gemini. Quale sia il migliore dipende quindi soprattutto dalle **preferenze** personali dell'utente.

Allora ecco una serie di spunti per iniziare a usare i chatbot IA da veri professionisti, anche per chi magari si è tenuto lontano da ChatGPT fino ad adesso. D'altronde, il report *State of the AI in business* lo dice chiaramente: sono i **“power user”**, coloro i quali hanno messo le mani sugli strumenti e ne hanno imparato i molti volti, che hanno le idee migliori su come applicare questa nuova tecnologia al contesto della propria azienda.

Nell'impiego degli LLM possono essere scelti diversi approcci, ognuno di loro fornisce risultati migliori a seconda del tipo di problema da affrontare. Ne analizziamo alcuni.

L' IA "Consulente"

Uno dei metodi più efficaci per utilizzare i chatbot IA, sul lavoro e nel tempo libero, è come **assistente virtuale specializzato** su di un qualsiasi tema di cui noi non siamo informati. In qualsiasi momento e in pochi secondi, l'IA può essere il meccanico che ci spiega come mai la moto non parte, il commercialista che ci illustra come accedere a un bonus o a una detrazione, l'esperto che ci consiglia come tutelarci nell'acquisto di un oggetto usato.

In particolare, uno studio ha mostrato che, sul posto di lavoro, l'IA può sostituire uno degli elementi di un team multidisciplinare, garantendo risultati finali soddisfacenti. D'altro canto, le ricerche mostrano anche che risultati eccellenti sono generalmente ottenuti se tutte le competenze necessarie sono apportate da umani. La consulenza di un'IA è dunque un **buon punto di partenza** ma, per prodotti sensibili, un professionista intelligente e coinvolto rimane insostituibile¹.

In pratica, per ottenere il massimo dall'IA, la chiave è assicurarsi che questa abbia **tutte le informazioni** che comunicheremmo al nostro "consulente". Per esempio, non ci dovremmo limitare a scrivere "come faccio a detrarre la spesa sui condizionatori?" ma dovremmo spronarlo a darci la risposta più completa possibile, aggiungendo alla domanda informazioni dettagliate: che vuoi installarli nella tua prima casa, che hai più di quarant'anni, che non hai aperto nessuna SCIA, che hai un lavoro dipendente a tempo determinato, che la tua casa è in classe energetica F, ecc.

¹

The Cybernetic Teammate: A Field Experiment on Generative AI Reshaping Teamwork and Expertise by Fabrizio Dell'Acqua, Charles Ayoubi, Hila Lifshitz-Assaf, Raffaella Sadun, Ethan R. Mollick, Lilach Mollick, Yi Han, Jeff Goldman, Hari Nair, Stew Taub, Karim R. Lakhani :: SSRN

Non diversamente da un professionista in carne ed ossa, l'IA lavora meglio quando la caliamo completamente nel problema, altrimenti scopriremo solo quanto è brava a dare risposte generiche, prolisse, e poco utili.

Potete sperimentare direttamente con ChatGPT (o un servizio analogo) quanto un **prompt** ricco di informazioni possa fare la differenza. Si chiama prompt il testo che inviate all'IA, che costituisce l'input dal quale il modello produrrà la propria risposta.

Provate a porre all'IA la seguente domanda:



Leggendo la risposta, quasi certamente vedremo che è ben scritta e di buon senso. In questo caso stiamo usando ChatGPT come tecnico IT, e come tale ci dirà sicuramente di provare a spegnere e riaccendere i nostri dispositivi, di controllare la password, ecc. Tutte indicazioni potenzialmente utili ma molto semplici, che probabilmente abbiamo già testato.

Proviamo adesso, invece, a spingere ChatGPT a darci una risposta più pertinente alle nostre condizioni, invece di limitarsi a raccomandazioni banali e frasi di circostanza. Ecco allora un prompt più adeguato:



Ho un problema di connessione Wi-Fi sul mio portatile aziendale, un Dell Latitude 7420 con Windows 11. Il problema si è presentato in seguito ad un aggiornamento del sistema operativo.

Il computer rileva correttamente la rete Wi-Fi aziendale, ma quando provo a connettermi, dopo qualche secondo mi dà il messaggio di errore: “Impossibile connettersi a questa rete”.

- Ho già riavviato il computer.
- Tutti gli altri dispositivi (smartphone, PC dei colleghi) sono connessi e funzionano.
- Non ho cambiato la password che fino a ieri funzionava, e che continua a funzionare sugli altri dispositivi
- Il mio portatile si connette correttamente all’hotspot del cellulare.



Ecco allora che la risposta dell’IA cambia completamente, e diventa molto più precisa e potenzialmente utile.



Glossario

Prompt: istruzioni digitate dall’utente nella finestra di dialogo del chatbot. Può avere l’aspetto di una finestra di conversazione di un’applicazione di messaggistica.

L' IA "Tirocinante"

Un altro possibile approccio all'uso dell'IA è quello che segue il modello del "tirocinante". Usando gli strumenti di IA generativa se ne imparano potenzialità e limiti, e a volte si può arrivare ad **assegnarle compiti** anche abbastanza complessi. L'IA diventa una sorta di stagista personale che fa parti di lavoro che puoi ricontrollare dopo che questa le ha fatte al posto tuo.

Ad esempio, con il prompt seguente, **Claude** (un'IA generativa analoga a Chat GTP) è capace di fare un'analisi abbastanza completa, seppur forse qualche specialista potrebbe trovarla carente in alcuni punti. Comunque, c'è da aspettarsi un risultato soddisfacente, completato nell'ordine di un minuto o due:



Voglio creare un portafoglio di ETF selezionati da diversi mercati mondiali, per un investimento a lungo termine e ad alto rischio (azionario):

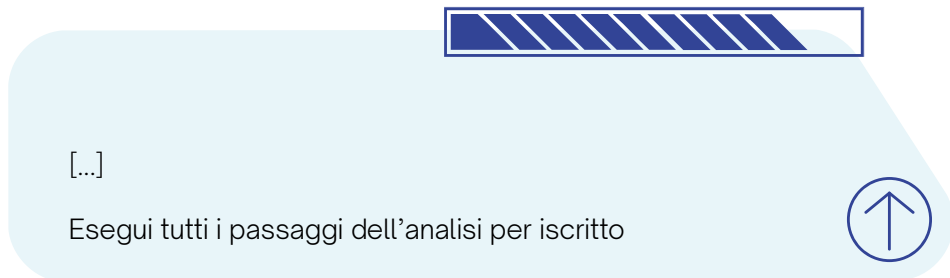
1. Identifica 5-10 mercati diversi in tutto il mondo, ragionevolmente grandi e separati (Stati Uniti, UE, Cina, Singapore, Brasile), cerca raccomandazioni finanziarie sul web e valuta selezioni diverse.
2. Trova alcuni ETF ben consolidati da ciascun mercato, adatte a investimenti pluriannuali.
3. Per ogni ETF trova i rendimenti storici e calcola il rendimento medio atteso. Trova inoltre la perdita massima a breve termine (drawdown - ad esempio controlla cosa successe nella settimana di esplosione del COVID) e fai il rapporto per calcolare l'indice RoMaD.
4. Armonizza i risultati per un confronto diretto e seleziona gli ETF con un indice più promettente. Fai anche un'analisi delle correlazioni tra questi ETF per selezionare i cinque più performanti.
5. Proponi un portafoglio che metta insieme questi ETF, pesati per performance, e confronta il risultato con un benchmark ragionevole e comparabile.

Esegui tutti i passaggi dell'analisi per iscritto e in sequenza.



In questo prompt sono da evidenziare le istruzioni precise e complete che mettono in chiaro che cosa ci si aspetta da un lavoro ben fatto. Stiamo chiedendo all'IA di fare delle analisi numeriche, di fare delle ricerche web, di fare delle valutazioni, e tutto ciò seguendo una sequenza precisa.

Un dettaglio di questo prompt potrebbe passare inosservato ma è fondamentale:



Questa frase è la chiave della tecnica chiamata **Chain of Thought** (CoT), scoperta e descritta in un paper del 2022, e considerata una pietra miliare nello sviluppo del **prompt engineering** e dell'IA generativa in generale.

Infatti, bisogna ricordare che quello che potrebbe sembrare una capacità di ragionamento, non lo è affatto. I modelli IA non sono e non vanno confusi per delle persone. Sono dei programmi che **elaborano un input e producono un output**. Non c'è garanzia che il testo prodotto da un'IA abbia un riscontro nella realtà, e non dobbiamo dimenticare, che un'IA non ha la capacità di pensiero, sta solo producendo un output verosimile in base agli input forniti. Ed è per questo che, inserendo nel prompt la richiesta di eseguire tutti i passaggi per iscritto, facciamo sì che l'IA effettivamente esegua i passaggi: perché **solo se li scrive li ha effettivamente fatti**.

Il metodo CoT è particolarmente versatile anche perché non è necessario includere a priori tutte le istruzioni nel prompt. Basta richiedere al chatbot di individuare i passaggi necessari per raggiungere un buon risultato e poi chiedere di seguirli e riportarli per iscritto. Tanto basta perché un'IA dia **risultati migliori** su qualsiasi problema di una certa complessità.

Nel 2024, i produttori di applicazioni IA hanno iniziato a includere questi elementi nei loro servizi: già GPT-4o scomponeva i quesiti

in passi e li eseguiva in chiaro anche senza un'istruzione specifica a riguardo. In seguito, Deepseek R1 contribuì a rendere popolare il cosiddetto reasoning, per cui i modelli producono una risposta “mentale” nascosta, prima di quella visibile e definitiva. In questa maniera, quando l'utente pone un quesito difficile da interpretare o particolarmente complesso, l'IA può rielaborarlo e scomporlo in passaggi più facili, senza dover produrre un testo completo con tutti i passaggi scritti esplicitamente.

Tutti i servizi di IA più importanti, come ChatGPT, Gemini e Claude, forniscono l'opzione di imporre al modello un periodo di reasoning prima di produrre la risposta. Questa opzione può essere utile quando non si vuole lavorare su prompt eccessivamente articolati ma si cerca comunque una risposta ben ponderata a un quesito difficile.



Glossario

Chain of Thought (CoT): metodo con cui un'intelligenza artificiale esplicita i passi che ha seguito per arrivare alla risposta finale di un quesito.

Prompt engineering: tecnica per formulare richieste in modo efficiente per ottenere risposte dai modelli di IA.

Reasoning: funzione di alcuni sistemi IA per collegare informazioni e farne valutazioni logiche.

L' IA web software

Pensare all'IA come a una persona, un “consulente”, o uno “stagista”, può essere utile come **modello mentale** quando ci avviciniamo a loro la prima volta, o impariamo cosa sono capaci di fare ma abbiamo già visto che è importante e prezioso imparare in cosa e perché le IA sono **diverse dagli umani**.

ChatGPT e simili, sono **software web**, e come tali andrebbero considerati. Non dobbiamo, con essi, auto-imporci convenzioni sociali umane, anzi ricordiamoci che, come con qualsiasi app, **siamo noi alla guida** dello scambio che abbiamo con essi.

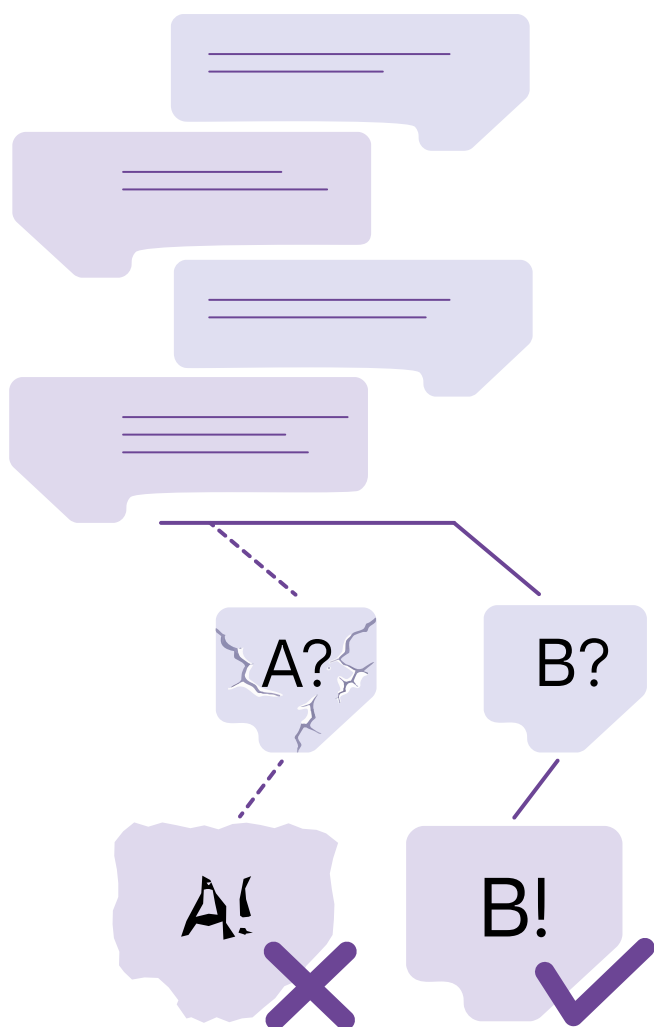
Anche l'interfaccia a chat, che è stata la chiave del successo di ChatGPT nel 2022, ci spinge, a trattare l'IA come se dall'altro lato fosse un umano a risponderci. Diventò virale ad aprile 2025 la notizia secondo cui gli utenti che ringraziano ChatGPT a fine conversazione costavano già decine di milioni di dollari ad OpenAI in consumi elettrici.

Conservare la consapevolezza che siamo noi a guidare la conversazione con il chatbot è uno strumento più potente di quanto sembri. È comune, infatti, tra i nuovi utenti delle IA generative, continuare una conversazione anche quando le risposte che stiamo ricevendo sembrano indicare che l'IA abbia **frainteso** le nostre richieste. Capita che una conversazione prenda una **piega inaspettata**, magari perché l'IA interpreta diversamente quello che abbiamo scritto, magari perché fa un errore ripetutamente e non capisce veramente il perché, o perché hai omesso un dettaglio fondamentale all'inizio. In questi casi è possibile tornare a un qualunque punto della conversazione e ricominciare lo scambio da lì. Con i servizi web come ChatGPT, puoi farlo. Una volta modificato un messaggio passato, l'IA genererà una nuova risposta da capo, più adeguata alla tua nuova richiesta.

In generale, l'ideale è **evitare conversazioni “fiume”** in cui si cerca di convincere l'IA di un suo errore ricorrente. Può avere i suoi vantaggi ma è più costoso e consuma più risorse. Piuttosto sono da preferire **conversazioni ad “albero”** che possono ramificarsi in

ogni punto in due, tre, fino a decine di direzioni diverse, ciascuna scaturita modificando un prompt, che genera in questa maniera una **nuova risposta** e un nuovo ramo della conversazione. Altrimenti, possiamo iniziare una nuova conversazione **da capo** ogni volta che vogliamo.

Quando interagiamo con l'IA, questa conosce solamente il contenuto della **conversazione visibile**, non sa nulla dei precedenti scambi, o delle varianti nascoste della presente conversazione. E funziona meglio così: che si concentri sullo scopo della conversazione attuale.



L'IA "ricercatore"

Dalla pubblicazione di ChatGPT, si è fin da subito discusso largamente di come avesse il potenziale per **sostituire i motori di ricerca web** classici, come Google. Secondo una ricerca di febbraio 2025, circa il 30% dei quesiti posti a ChatGPT è assimilabile a una ricerca web classica¹. Per quanto triviale, si tratta di un caso d'uso talmente comune che i motori di ricerca stessi (Google, Bing) stanno **integrando** dei testi generati da IA nella schermata dei risultati.

Si può confrontare l'approccio classico alla ricerca web con quello tramite IA generativa, e vedere come questa risulta più veloce e immediata, andando a pescare le informazioni di interesse all'interno dei siti web, elaborandole e riportandole direttamente come richieste. Il rovescio della medaglia è che in questo modo l'utente **perde il controllo** sulle fonti consultate e sul loro contesto. Si può compensare, in parte, richiedendo di **citare le fonti** e fornendo i link corrispondenti a ogni informazione.

Questa unione tra IA generativa e ricerca delle fonti viene indicata tecnicamente con la sigla RAG (**Retrieval Augmented Generation**), firmata dai ricercatori di Meta in un articolo del 2020, ed è un tema intensamente ricercato. Permette all'IA di generare risposte concretamente **basate su fonti**, invece di affidarsi solamente alla meno affidabile conoscenza interna del modello ma può in alcuni casi peggiorare le risposte generate se le fonti su cui si basa la risposta sono irrilevanti, errate, o confuse.

Nel concreto, possiamo immaginare alcune situazioni in cui la conoscenza del modello, pur poco affidabile, può arrivare a una risposta migliore rispetto a una ricerca web. Possiamo sperimentare la differenza sia con ChatGPT, che permette di scegliere tra una modalità automatica o una in cui la ricerca web è sempre attiva, sia con Claude che ci lascia disattivare completamente la ricerca web. Le risposte nei due casi (con ricerca web attiva o meno) possono essere sostanzialmente diverse, ad esempio, con un prompt come il seguente:

¹ Investigating ChatGPT Search: Insights from 80 Million Clickstream Records



Quale autore greco scrisse qualcosa sulla falsariga di "è dura essere comandati da chi ci è inferiore"?



La ricerca web produce pochi risultati rilevanti, mentre il modello, basandosi solamente sulla propria memoria interna, suggerisce una risposta unica, che però va assolutamente **verificata** se non è supportata da fonti.



Glossario

Retrieval Augmented Generation: la RAG è una tecnica in cui l'AI recupera informazioni esterne aggiornate per creare risposte più precise, affidabili, comprensibili all'utente.

Il metodo RAG, comunque, non è affatto limitato alle ricerche web: si può applicare a qualsiasi dominio della conoscenza/corpus di documenti. Ecco alcuni esempi di strumenti RAG applicabili a domini diversi:

- ◆ **NotebookLM** è un servizio web di Google nel quale è possibile creare **knowledge domain** personali (chiamati notebook), nel quale inserire un certo numero di documenti di testo (ad esempio in formato pdf), e poi interrogare l'IA su questioni trattate all'interno dei documenti caricati. NotebookLM include anche una serie di funzionalità aggiuntive, come la mappa mentale o l'infografica, che sintetizzano visivamente i temi trattati all'interno di tutti i documenti caricati.

- ◆ **Copilot** di Microsoft, integrato nella suite Office per imprese, permette in pochi semplici passaggi, di creare dei chatbot integrati su Teams, che usano come base di conoscenza una cartella Sharepoint di documenti aziendali. Il chatbot può poi essere condiviso con gli altri dipendenti. Esistono altri servizi simili rivolti agli utenti di ChatGPT (i cosiddetti **GPT**) o Gemini (chiamati **Gem**).
- ◆ **Elicit e Scite** sono altri esempi di servizi RAG ma basati su letteratura scientifica e ottimizzati in particolare per le metanalisi. Elicit, ad esempio, permette di creare “notebook”, cioè chat basate su degli articoli scientifici trovati dall’IA, oppure “review”, in cui l’IA trova e applica automaticamente una serie di filtri agli articoli scientifici raccolti, selezionando solo quelli rilevanti, per poi estrarre e sintetizzare i dati riportati negli articoli.
- ◆ **Deep Research** è una funzionalità che può essere attivata su Gemini e ChatGPT, ed è qualcosa di più elaborato di un semplice sistema RAG: se la attiviamo vediamo che il sistema ci chiederà di definire precisamente il tema della ricerca (in particolare Gemini presenta una finestra apposita), prima di avviare una ricerca, che può necessitare anche di 10-15 minuti. Il sistema scompone la richiesta in capitoli, esegue ricerche multiple, cercando di coprire tutte le tematiche, collezionando decine, o, a seconda del tema, anche centinaia di fonti, e infine redige un report strutturato con i risultati della ricerca. Gemini, in particolare, arriva a comporre documenti di molte decine di pagine, seppur con una tendenza alla ripetizione. Il risultato è un’analisi abbastanza completa su di un tema di cui magari potrebbe mancare un testo comprensivo.

L'IA "agente"

Siamo arrivati, a questo punto, agli strumenti più avanzati basati su IA, chiamati **agenti**. Possiamo immaginare un agente IA come un **copilota**, che ha accesso ai nostri stessi strumenti (o perlomeno quelli a cui gli concediamo l'accesso) e, se glielo chiediamo, può sostituirci in tutto e per tutto nello svolgimento di alcuni compiti.

In particolare, è possibile avere un agente IA che compie azioni direttamente sul nostro computer, inviando comandi testuali tramite l'interfaccia a **riga di comando**. Questo tipo di interfaccia è particolarmente nota ai programmatori, sostituita per gli utenti comuni da cartelle e finestre ma, ancora adesso, rimane uno strumento estremamente versatile e potente.

Non è un caso che molti dei software basati su agenti IA siano disegnati per programmatori: ad esempio **Copilot per VSCode**, **Claude Code**, **Google Antigravity**, sono tutti strumenti che usano l'IA per generare codice di programmazione, modificare programmi esistenti, testarli ed eseguirli, con un alto grado di autonomia.

È comunque possibile, con un po' di fantasia, utilizzare queste applicazioni per compiti diversi dalla programmazione, ad esempio per organizzare dei documenti, elaborare file e creare riassunti, anche in serie, eseguire lavori che richiedono molti passaggi e controlli (gli agenti sono ottimizzati per lavorare sullo stesso compito anche molto a lungo).

Altrimenti, sono disponibili agenti online come **Manus**, che permette di collegare all'IA i propri account di posta elettronica e calendario, browser e varie applicazioni, soprattutto nel campo del lavoro/produktività. In seguito al successo di Manus, pubblicato a marzo 2025, ChatGPT e Claude hanno entrambi integrato lo stesso tipo di connettori ad applicazioni esterne, e possono, su richiesta o necessità, estrarre dati dalle applicazioni connesse, o anche eseguire azioni su quelle app. Ad esempio, si può interrogare l'agente con questioni del tipo:



Qual è il numero di telefono del PM del progetto G.A.I.A.?



Il che ci potrebbe evitare di ricercare l'informazione nella casella di posta elettronica.

Le possibilità sono molto vaste, anche se a volte di scarsa utilità per l'utente comune. Non di meno, in alcune situazioni particolari, un agente IA potrebbe rivelarsi la soluzione perfetta per accelerare il lavoro su di un compito specifico. Sperimentando con questo tipo di soluzioni di frontiera, potremmo anche renderci conto che nel nostro quotidiano c'è margine per **delegare flussi di lavoro** significativi ad applicazioni basate su agenti IA.

Tutti i pesci del mare

Abbiamo menzionato, fino a qui, già numerose IA differenti che utilizzano ciascuno una varietà di modelli. I modelli di **OpenAI** sono lo standard di fatto e **ChatGPT** è uno dei servizi più versatili e capaci. **Google** si è nel tempo guadagnato un rispetto analogo per i propri modelli della famiglia **Gemini** e anche **Nano Banana** merita una menzione per la velocità di elaborazione e la sua qualità. Poi c'è **Grok**, sviluppato da **xAI**, che, all'opposto, si distingue per il tono informale.

Queste sono le famiglie di modelli più avanzate disponibili al momento in cui scriviamo. Hanno tutte in comune il fatto di essere sviluppate negli USA e utilizzabili solo tramite servizi Cloud, per preservare il segreto aziendale sul contenuto algoritmico dei modelli. La scelta però è più ampia di quanto sembri. Tra i più famosi model-

li di libero accesso (il cosiddetto Open Source) ci sono quelli appartenenti alle due famiglie **Deepseek e Kimi**, pubblicati dalle due aziende cinesi Deepseek e Moonshot AI. Sono modelli che possono essere copiati su un server di nostra gestione e usati liberamente in locale, senza nessun costo, e sono entrambi fortemente ottimizzati per risparmiare risorse computazionali durante l'utilizzo, pur garantendo performance di tutto rispetto.

Infine, anche l'Europa ha il suo modello IA sviluppato da **Mistral** l'azienda leader europea nel settore. Anche Mistral, peraltro, ha pubblicato alcuni suoi modelli Open Source che possono essere scaricati e usati in locale.



Glossario

Agente: sistema di intelligenza artificiale che può osservare informazioni, prendere decisioni autonome ed eseguire azioni per raggiungere obiettivi specifici.

Riga di comando: interfaccia testuale dove l'utente scrive comandi in un linguaggio di programmazione prestabilito, permettendo al computer di eseguire rapidamente operazioni precise.

Qualche spunto ulteriore

In questa guida ci siamo sforzati di creare un percorso che, anche partendo da zero, potesse insegnare a tutti come avvicinarsi all'uso delle IA. Per imparare a usarle in maniera efficiente e, allo stesso tempo, conoscerle un po' meglio comprendendo le loro potenzialità. Per questo ci siamo concentrati solo su alcuni aspetti che ci hanno portato a esplorare le funzionalità più versatili e interessanti dell'IA generativa. Tuttavia, c'è molto da scoprire se si continua ad esplorare.

Multimodalità

Per esempio, non abbiamo ancora parlato delle **funzionalità multimodali**, come l'interfaccia vocale, o la comprensione e generazione d'immagini. La multimodalità delle IA è stata considerata uno degli aspetti più interessanti tra 2024 e inizio 2025, quando furono per esempio pubblicati i modelli GPT-4o e Llama 4. Nella pratica, i nuovi modelli di IA generativa tendono a specializzarsi su uno tra testo, immagini o video. Ad esempio, i modelli Gemini 3 (testuale) e Nano Banana 2 (immagini) sono separati nonostante Google li abbia sviluppati e pubblicati nello stesso periodo.

Naturalmente, rimangono due modelli fortemente integrati, e molto utili se si lavora con dati multimodali. Per fare un esempio, Gemini e simili sono particolarmente efficaci nel comprendere testo presente nelle immagini e a contestualizzarlo, oppure analizzare simboli, e sono usati per generare immagini alternative a quelle stock.

Documenti

Gemini utilizza Nano Banana anche per creare presentazioni o singole diapositive in Google Slides. Questo è un approccio diverso da quello di Claude, il quale ha una funzionalità per creare presentazioni creando layout composti di testo ed immagini.

In generale, non tutti i servizi di IA sono ugualmente integrati con **documenti di formati differenti**. La maggior parte dei servizi, per esempio, permette di allegare ai propri prompt documenti in pdf ma non tutti accettano file Word. Claude poi, è particolarmente efficace con i fogli di calcolo.

L'arte del prompting

In questa guida abbiamo ancora parlato poco di **prompting**, ovvero **come scrivere il comando** perfetto, per cambiare i risultati del modello e ottenere esattamente ciò che si desidera. Non perché il prompt non sia importante – può fare un'enorme differenza – ma semmai perché è un tema che già gode di grande attenzione anche in ambienti poco specialistici. In linea di principio, il prompt perfetto dipende moltissimo sia dall'obiettivo, sia dal modello. Con la pubblicazione di una nuova versione del modello, tutte le raccomandazioni di prompting possono diventare rapidamente obsolete.

Di conseguenza, è sempre utile fare la propria **personale esperienza** su come interagire con i vari modelli e scoprire in autonomia cosa funziona meglio per noi. Ecco però alcune idee, se volete provare a variare il vostro prompt.

- ◆ Specifica **come** verrà utilizzato il testo generato e **a chi** è rivolto: il modello enfatizzerà le tematiche di maggiore interesse per i destinatari, adattandosi alla loro cultura e alle conoscenze tecniche del pubblico. Una declinazione particolarmente utile di questa idea è quando si usa l'IA per **studiare nuovi temi**: dichiarare il proprio background permetterà l'IA a creare una versione personalizzata della lezione che si vuole apprendere.
- ◆ Il **persona prompting** consiste nel dare a un modello un'**identità** rilevante al tema della richiesta. Ad esempio, possiamo scrivere "Sei un consulente finanziario esperto e imparziale..." prima di chiedere in che modo finanziare l'acquisto della nostra prossima auto. Il persona prompting cambierà l'approccio della rispo-

sta e il tono. Ci sono anche ricerche che hanno mostrato come il persona prompting può spingere il modello a dare risposte fattualmente più corrette anche se ormai i modelli più recenti tendenzialmente non ne hanno più bisogno.¹

- ◆ La **lingua** in cui scriviamo il prompt può influenzare il risultato. L'**inglese** è la lingua più diffusa per l'addestramento dei modelli IA, quindi, può valere la pena di usarla anche nelle nostre ricerche. In generale, se un tema è affrontato con approcci diversi da diverse comunità linguistiche, la risposta ne sarà influenzata.
- ◆ È possibile utilizzare alcuni semplici **stragemmi grafici** per guidare l'attenzione dell'IA. Questi strumenti servono a porre l'accento su parti specifiche del contenuto del prompt, indicando al modello la gerarchia delle istruzioni. Il **maiuscolo**, ad esempio, viene interpretato come un segnale di importanza critica ed è lo strumento più efficace per comandi imperativi o vincoli che non devono assolutamente essere ignorati. Per strutturare prompt più lunghi si possono creare **titoli e sottotitoli**, ponendo un certo numero di cancelletti a inizio della riga titolo (più cancelletti significa sottotitoli gerarchicamente inferiori, un cancelletto unico si riserva al titolo generale di un documento). Una serie di trattini "----" indica una separazione tra sezioni del prompt. Allo stesso modo si possono usare anche tag Xml (ad esempio "<citazione>...</citazione>"). Usare questi indicatori aiuta l'IA a comprendere meglio la struttura e le priorità all'interno del prompt, aumentando le possibilità di ricevere una risposta adeguata.
- ◆ Per rendere l'output dell'IA immediatamente utilizzabile e risparmiare tempo sulla rielaborazione manuale, è estremamente utile **definire in modo esplicito il formato** della risposta. Se i dati devono essere implementati in un foglio di calcolo o in una presentazione si può chiedere all'IA di strutturare il testo di conseguenza: "Considera il seguente testo: [...]. Correggi gli eventuali errori di grammatica e punteggiatura e restituisci il testo corretto e già formattato in **Markdown** in modo da copiarlo e incollarlo direttamente in un documento Word."

¹ Prompting Science Report 4: Playing Pretend: Expert Personas Don't Improve Factual Accuracy by Savir Basil, Ina Shapiro, Dan Shapiro, Ethan R. Mollick, Lilach Mollick, Lennart Meincke :: SSRN

Attenti alla privacy!

Perché le IA ci sappiano assistere efficacemente, può capitare che sia necessario condividere con loro informazioni dettagliate sul lavoro di cui ci stiamo occupando, informazioni che però potrebbero essere protette da **segreto aziendale** o da **accordi di riservatezza**. Condividere questi dati con l'IA può essere problematico.

Naturalmente, tutti i servizi basati su IA che operano in Unione Europea sono tenuti a **rispettare il GDPR**: il regolamento comunitario sulla privacy. Perciò il primo passo è verificare che il chatbot che stiamo usando sia rivolto ad utenti europei. Essendo tutelati dal GDPR abbiamo la possibilità di richiedere in qualsiasi momento la **cancellazione dei dati** che abbiamo condiviso col chatbot.

Il GDPR, inoltre, proibisce che il fornitore del servizio IA utilizzi i dati al di fuori della **correttezza** (fairness) e del **legittimo interesse**, e li venda, ad esempio, ai nostri competitor. Tuttavia, c'è un altro utilizzo dei nostri dati che può costituire un problema per la loro riservatezza: quasi tutti i servizi IA utilizzano i dati per **aggiornare** i modelli di Intelligenza Artificiale, attraverso un addestramento che potrebbe essere basato su ciò che noi gli stiamo scrivendo.

Per questa ragione, è importante condividere informazioni riservate solamente con IA che **non utilizzano** questi dati per l'**addestramento dei modelli**. Attenzione! Sia **ChatGPT** che **Claude** e **Gemini** utilizzano i nostri dati con questa finalità. Per proteggerne la confidenzialità dobbiamo modificare le impostazioni predefinite del servizio.

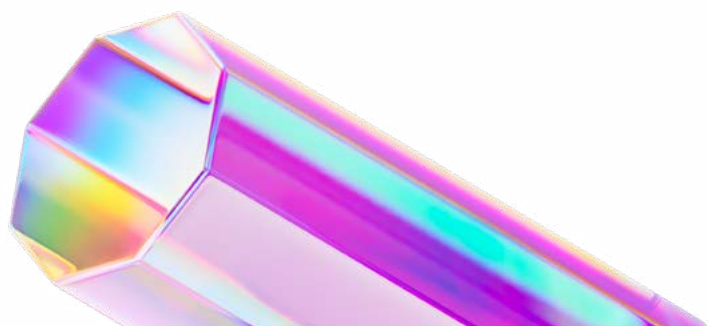
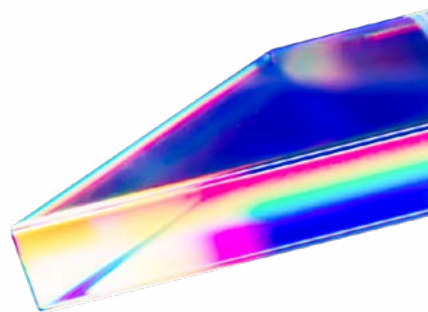
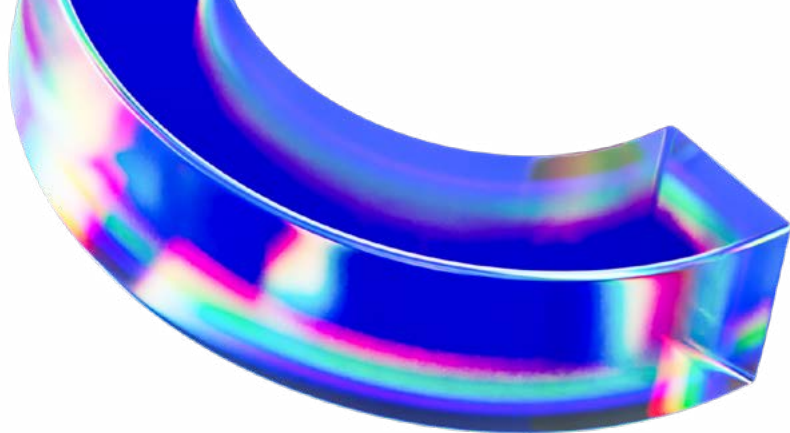
Gemini, in particolare, non consente agli account gratuiti di modificare questa impostazione. Quando invece è utilizzato tramite **Google Workspace**, essendo destinato a imprese e professionisti, gli input degli utenti non vengono mai utilizzati per l'addestramento dei modelli.

03

┌ L'IA generativa come sistema aziendale

Da strumento personale
a sistema autonomo





Fin qui abbiamo parlato di come affidarsi alle IA per il lavoro e le sfide quotidiane, all'interno della nostra sfera personale in cui l'IA è assistente e strumento da noi direttamente controllato. Un'innovazione di tale portata però sembra quasi aspettare soltanto di essere colta da ogni gestore di impresa per farne valore economico e ascensore di competitività. Proposito auspicabile ma che nel concreto fa molta fatica a essere messo in pratica.

Il report *State of AI in Business* del MIT è il più approfondito studio fatto finora sull'**adozione della tecnologia per l'innovazione aziendale**. Questo documento affronta alcuni dei falsi miti che influenzano l'adozione di IA in azienda, sia positivamente che negativamente. Eccone alcuni:

<< L'Intelligenza Artificiale sostituirà la maggior parte dei posti di lavoro nei prossimi anni >>

✦ Dallo studio emerge che l'impiego di Gen IA in azienda non ha portato a un aumento significativo dei licenziamenti. I licenziamenti sono stati limitati solo a settori già significativamente impattati dall'impiego di IA. In generale, ci sono opinioni molto discordanti tra i dirigenti delle aziende interpellate in merito all'impatto che avranno le IA sui posti di lavoro nei prossimi 5 anni.

<< L'Intelligenza Artificiale generativa sta trasformando le imprese >>

✦ L'adozione è elevata, ma la trasformazione è rara. Solo il 5% delle aziende ha strumenti di Intelligenza Artificiale integrati nei flussi di lavoro su larga scala e 7 settori su 9 non mostrano alcun reale cambiamento strutturale.

<< Le aziende sono lente nell'adottare nuove tecnologie >>

✦ *Le aziende sono ansiose di adottare l'Intelligenza Artificiale e il 90% ha seriamente valutato l'acquisto di una soluzione IA.*

<< I principali ostacoli all'Intelligenza Artificiale sono: la qualità del modello, aspetti legali, politiche sui dati e di rischio >>

- ✦ Ciò che la frena davvero è che la maggior parte degli strumenti di Intelligenza Artificiale non apprende e non si integra bene nei flussi di lavoro.

<< Le aziende migliori stanno sviluppando i propri strumenti >>

- ✦ I progetti interni falliscono con il doppio della frequenza di quelli basati su strumenti già disponibili.

A fianco di un dato certamente scoraggiante sul tasso di successo dei progetti GenAI, il report del MIT nota che, in caso di successo, i risultati si ottengono molto rapidamente e che i prodotti IA, così sviluppati, incontrano subito un grande interesse del mercato.

Le imprese stanno quindi dando molta importanza a questa tecnologia, e sono decisamente interessate alle potenzialità di sviluppo che porta con sé. Tenendo in conto tutte le considerazioni fatte fino ad ora, è importante capire che tipo di progettualità basata su AI generativa ha maggiori **probabilità di successo** e capire quali applicazioni sia più conveniente sviluppare.

L'IA generativa, che siamo così abituati a vedere sotto forma di chatbot, viene applicata spesso in progetti di marketing ed esperienza utente, anche grazie al fatto che forniscono **risultati misurabili** che ne scaturiscono (user retention, volume di presentazioni del prodotto...) e sono direttamente allineati con le richieste del management. Questi prodotti hanno spesso applicazioni vaste e non ben definite e non rappresentano un vero guadagno rispetto a strumento **generalista** come ChatGPT o Claude.

Lavoratori e professionisti si sono dimostrati estremamente permeabili all'avvento dei chatbot IA e, essendo abituati ad usarli nel quotidiano, li privilegiano a prodotti sviluppati ex-novo che risultano **meno familiari** e non altrettanto flessibili, anche se realizzati utilizzando le stesse tecnologie e modelli. Di fatto, l'utente per il suo

lavoro preferisce spesso sottoporre lo stesso problema a ChatGPT piuttosto che a uno strumento fatto apposta allo scopo.

Dalle parole degli specialisti, intervistati nello studio del MIT sopracitato, queste limitazioni emergono ancora più chiaramente.

<< Abbiamo visto decine di demo quest'anno. Forse una o due sono effettivamente utili. Le altre sono solo wrapper o progetti di ricerca. >>

<< Lo strumento di Intelligenza Artificiale che abbiamo acquistato forniva riepiloghi rigidi con opzioni di personalizzazione limitate. Con ChatGPT, posso guidare la conversazione e iterare finché non ottengo esattamente ciò di cui ho bisogno. La differenza fondamentale in termini di qualità è notevole: ChatGPT produce costantemente risultati migliori, anche se il nostro fornitore afferma di utilizzare la stessa tecnologia di base. >>

Identikit dello strumento IA di successo

Dallo stesso report da cui sono tratti questi dati e citazioni, possiamo anche ricavare una mappa di quali progetti sono invece riusciti ad apportare miglioramenti stabili in termini di produttività o bilancio in azienda.

Da un lato, ci sono prodotti semplici che svolgono o accelerano, in serie e su larga scala, task ben definiti e di bassa complessità (revisioni di contratti, categorizzazione della spesa, contenuti personalizzati per campagne di comunicazione, ecc.). Questi sono i cosiddetti **Fast Wins**.

Al di là dei prototipi più semplici, i prodotti IA di successo devono superare i limiti dei chatbot generalisti per apportare un valore reale. In particolare, questi dovrebbero includere:

- ◆ **Forte allineamento con il processo aziendale interessato:** il prodotto IA è realmente utile se sviluppato con una compren-

sione approfondita dei processi che deve integrare.

- ✦ **Integrazione con gli strumenti e sistemi aziendali:** un prodotto isolato dai dati e dagli strumenti aziendali aggiornati non è migliore di ChatGPT.
- ✦ **Flessibilità e adattabilità ai cambiamenti:** il prodotto IA deve essere in grado di evolvere insieme ai processi aziendali per restare utile e rilevante.
- ✦ **Capacità di migliorare nel tempo:** occorre che lo strumento sappia imparare e migliorarsi per non compiere sempre lo stesso errore.

Per realizzare questi prodotti di successo, le imprese si affidano tipicamente start-up e fornitori esterni, trattati come partner strategici. Questo approccio si avvicina al **Business Process Outsourcing (BPO)**, e consente una profonda personalizzazione delle soluzioni e un'integrazione efficace nei processi aziendali, grazie a un continuo adattamento basato sul feedback degli utenti e assicurato dal controllo continuativo da parte dell'azienda committente.

Altro fattore di successo per i progetti di IA generativa è la **proattività degli utenti** e dei lavoratori che conoscono la tecnologia essendo già abituati ad utilizzarla. Le iniziative calate dall'alto, affidate a laboratori IT esterni o centralizzati hanno una frequenza di successo molto inferiore rispetto a situazioni in cui i lavoratori stessi sono parte dello sviluppo. Grazie alla collaborazione degli utenti direttamente interessati, le soluzioni vengono sviluppate molto più velocemente e risultano più efficaci e maggiormente allineate ai processi aziendali.

Nel pratico, il processo mentale dietro allo sviluppo di un progetto di IA, può essere quello di applicare a un prompt il concetto di **scalabilità**, automatizzando parte dell'input che varia in base al contesto aziendale e al compito in questione. L'obiettivo diventa quindi di ottenere un **output affidabile** e immediatamente integrabile nei sistemi aziendali.

Questi prodotti possono migliorare drammaticamente la produttività in processi di backoffice, poco visibili e difficilmente misurabili. In questi casi, l'azienda può arrivare a permettersi di tagliare sulle spese di consulenza e BPO sostituendoli con i prodotti di IA generativa.

Parole chiave per l'IA generativa fatta sistema

Dopo aver letto tutta questa guida, potreste avere voglia di utilizzare l'IA generativa all'interno della vostra azienda. Avete considerato i limiti e i vantaggi, e vi siete convinti che l'IA si possa accomodare al cuore dei vostri sistemi aziendali, accelerandoli e potenzialmente aumentandone grandemente la portata. Magari avete già in mente un partner o una start-up con cui collaborare per dare luogo a questa transizione. Cosa aspettarsi adesso?

Se siete già utenti abituali di IA e chatbot, e avete già preso in mano gli strumenti di cui parliamo nella prima parte di questa guida, allora avete già chiara la situazione. Adesso però costruiamo insieme il vocabolario per parlare della parte nascosta, più tecnica, ma fondamentale per affrontare il discorso dell'IA come sistema integrato in azienda. Ecco un glossario essenziale:

API e Foundation Model

Quando scriviamo e conversiamo con ChatGPT, stiamo utilizzando un'applicazione web, una semplice interfaccia per ciò che sta sotto: il modello, anzi i modelli, di linguaggio, o LLM. ChatGPT stesso, in quanto interfaccia, **sceglie dinamicamente** il modello da utilizzare in base all'input utente (esiste un modello IA specifico anche solo per fare questa scelta), e poi prende l'intera conversazione e la dà in pasto al modello generativo scelto, che da tale input produce l'output che noi finalmente leggiamo.

ChatGPT e Claude, e le altre interfacce che troviamo sul web, sono servizi rivolti direttamente agli utenti. Quando dobbiamo creare un servizio particolare per la nostra azienda, dobbiamo passare dalle **API (Application Programming Interface)** di OpenAI o Anthropic in quanto Cloud Provider di modelli IA. In particolare, queste aziende, ma anche altre tra cui i maggiori provider cloud, come Azure, Google e AWS, danno accesso ai **foundation model**, gli LLM fondamentali, mattoni elementari delle applicazioni basate su IA generativa. La sigla API spiega cosa c'è dietro: un programmatore può sviluppare la propria applicazione, ovunque sia installata, per collegarsi via web ai cloud provider. L'applicazione invia gli input da elaborare da remoto ai data center di Google e OpenAI, Microsoft o Amazon, e si fa restituire gli output.



Glossario

API: interfaccia che permette a software diversi di comunicare e scambiarsi dati automaticamente.

Foundation Model: modello di intelligenza artificiale molto grande e versatile, addestrato su tanti dati per svolgere molti compiti diversi.

One-shot learning: tecnica in cui l'IA impara a riconoscere qualcosa da un solo esempio invece che da molti.

Edge: tecnologia che elabora i dati vicino al luogo in cui vengono prodotti, riducendo la dipendenza da internet.

Cloud: server remoti che forniscono potenza di calcolo e spazio, accessibili tramite internet per elaborare e conservare dati.

One-shot Learning

Con **one-shot learning** (anche chiamato **zero-shot learning**) si intende la possibilità di dare a un LLM le istruzioni per produrre un determinato output da un certo input variabile. Il paragone, in questo caso, è quello del **machine learning** tradizionale, in cui si utilizzano vaste raccolte dati per addestrare un modello da zero (o quantomeno parzialmente) a questo preciso scopo: prendere un input di qualche tipo (immagine, testo, documento, ecc.) ed estrarne un sentiment, dei dati particolari, una categorizzazione, o altro.

Gli LLM ci permettono di saltare completamente la fase di addestramento del modello e di far produrre gli output direttamente ai foundation model, premesso un prompt adeguato.

Tramite one-shot learning possiamo affidare all'IA compiti tediosi come la classificazione di centinaia o migliaia di documenti o supporti multimediali, per categorie o punteggi, l'estrazione di informazioni strutturate o la rielaborazione degli input.

Allo stesso modo, lo one-shot learning può essere un ottimo alleato per la **prototipazione** veloce di applicazioni basate su IA tradizionali, iniziando a validare la soluzione ancor prima dell'addestramento di un modello di machine learning.

Bisogna comunque tener conto che gli LLM non sono capaci di sostituire tutti i tipi di machine learning tradizionale, e che alcuni casi esemplari, come la regressione di dati tabellari, non sono alla portata del one-shot learning. Inoltre, anche dove questo funziona bene, è tendenzialmente più **lento** di una soluzione sviluppata ad hoc, e comporta **spese continue** per l'utilizzo. Il machine learning rimane la soluzione migliore sul lungo termine.

LLM on the edge

Quando invece è specificamente richiesto un LLM, esiste comunque la possibilità di avviare il modello su un dispositivo **locale** (in edge) **anziché in cloud**. Solo i modelli più minuti possono girare su

un computer portatile o su un cellulare, mentre modelli di medie e grandi dimensioni hanno bisogno di server con hardware particolarmente potente, soprattutto se le risposte devono essere elaborate in tempi brevi.

Le ragioni principali per preferire l'edge al cloud, in questo caso, sono la possibilità di funzionare indipendentemente dalla connessione remota e la garanzia totale di **privacy**, che può essere particolarmente cara a qualche stakeholder: una cosa è un'assicurazione legale di riservatezza, altra è la sicurezza materiale che i dati non escano mai dal sistema.

I modelli disponibili per essere eseguiti sull'edge sono quelli **open-source**, che non richiedono nessun costo per essere installati e utilizzati, ma che non sono esattamente al livello dei modelli di frontiera come quelli di OpenAI e Anthropic, che sono molto attente a non condividerli.

RAG

Questa funzione è già stata introdotta nei capitoli precedenti per parlare di IA come ricercatore. Nel contesto aziendale può rivelarsi utile perché fornisce dinamicamente all'IA informazioni essenziali sul contesto aziendale in cui questa opera. Tuttavia, è importante avere **aspettative realistiche** sull'efficacia dell'algoritmo che estrae i dati dalla **Knowledge Base (KB)**. Per risultati ottimali, infatti, occorre una valutazione attenta dei documenti di cui la KB si compone. A volte è necessario fare una selezione di questi contenuti e poi un'ottimizzazione dell'algoritmo di recupero basata sui documenti raccolti e sull'uso previsto. È facile, infatti dare per scontate performance al livello di quelle di Google, dimenticando che Google è la soluzione a un problema algoritmicamente complesso, e che non a caso è diventata una compagnia multimiliardaria. Il pacchetto Office, rende possibile creare dei semplici chatbot potenziati da RAG anche utilizzando Copilot, e individuando la KB tra i documenti di SharePoint.

Prompt Engineering

Il prompt engineering è un po' l'arte e un po' la scienza di scrivere un buon messaggio all'intelligenza artificiale, per "convincerla" a eseguire un certo compito per tuo conto. Si tratta di fornire all'IA tutte le informazioni utili al suo svolgimento, assicurandosi che in condizioni variabili queste siano presentate in maniera ottimale e senza lacune che possano ad esempio causare allucinazioni da parte dell'IA. Si tratta anche di una disciplina che, tramite procedure di qualità, assicura che i risultati ottenuti corrispondano alle aspettative.

È facile sentir dire che il prompt engineering è una formula magica di incredibile potere, o al contrario che sia un'inutile cortina di fumo senza consistenza. La realtà è che un lavoro approfondito sui prompt può essere consigliabile o meno a seconda del caso: per applicazioni con uno scopo ben definito, prompt ben affinati possono conseguire risultati impossibili da raggiungere se non con metodi molto più costosi.



Glossario

Open source: software con codice visibile e modificabile da tutti, sviluppato in modo collaborativo e libero da licenze proprietarie.

Knowledge Base: raccolta organizzata di informazioni e documenti che funge da sorgente di risposte e soluzioni.

Prompt chaining: tecnica che collega più richieste successive a un'IA per ottenere risultati complessi in modo graduale e controllato.

Sistemi multi-agente

Nell'ambito del prompt engineering esistono vari metodi che permettono di usare non un solo prompt ma diversi, in alcuni casi usando in sequenza (**prompt chaining**), oppure organizzando schemi di flusso che usino prima un prompt e poi un altro per elaborare un certo input, secondo delle regole applicate sempre da IA. Anche questo è un modo per massimizzare l'efficacia dell'IA su problemi specifici tramite prompt ben organizzati.

In un sistema come questo, l'IA che risponde a ciascun prompt viene chiamata **agente** perché ha un compito specifico all'interno del sistema. Ad esempio, vi può essere un agente che fa da **pianificatore** e un agente che fa da **esecutore**, un altro da **critico** che valuta il lavoro dell'esecutore, ecc.

IA agentica

Per aggiungere confusione, nel gergo IA, la parola agente si è affermata a un certo punto con un significato diverso, ben distinto da quello che abbiamo appena visto: è un agente un'IA capace non solamente di generare risposte, generalmente testuali, dagli input, ma anche di **interagire** con il mondo attraverso una serie di strumenti che siano stati messi a sua disposizione.

L'interfaccia AI-strumenti si indica con l'espressione **tool usage**, o talvolta **function calling**, e tipicamente sono implementati in sinergia con il **reasoning** per cui l'agente può decidere di usare uno strumento, ricevere da questo strumento un risultato, ragionarci su e prendere a quel punto una nuova decisione, come ad esempio usare un altro strumento, scrivere una risposta testuale verso l'utente o riutilizzare lo stesso strumento.

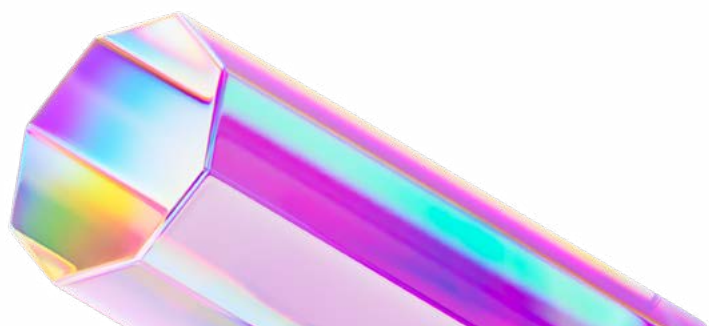
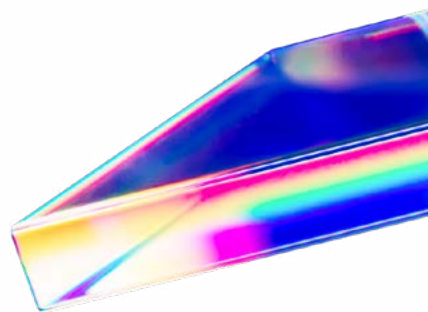
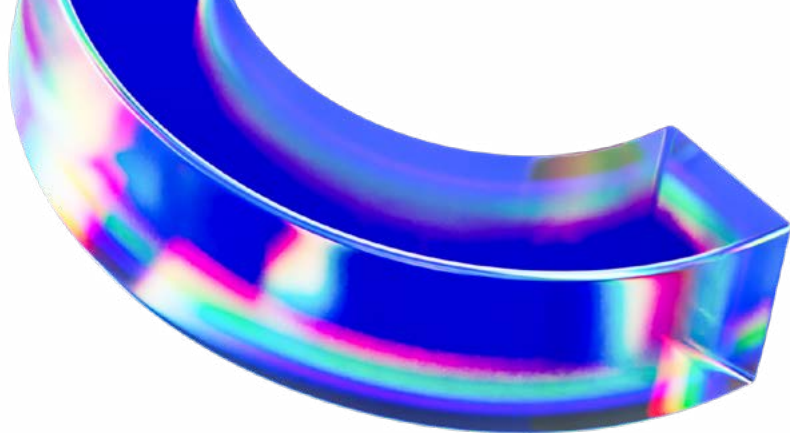
L'IA può **ripetere** questo ciclo strumento-reasoning fino a che non giudica di poter dare una **risposta completa** all'utente. L'IA adotta in questo modo un flusso di lavoro autonomo ritornando sui propri passi, senza intervento da parte dell'utente, dopo essere incappati in un vicolo cieco. L'IA può così svolgere compiti di una certa complessità in totale autonomia.

L'IA agentica è stata oggetto di importanti **investimenti** ultimamente e si può constatare come i modelli di frontiera dei principali provider siano stati sviluppati con questo tipo di applicazioni in mente. Non solo sono ottimizzati per un tool usage disinvolto ma anche per continuare a lavorare sulla stessa richiesta per molto tempo **senza fermarsi**. Ad esempio, GPT-5.2 può continuare a lavorare **per ore** sullo stesso compito, mentre GPT-5.1 si arrende entro mezz'ora o poco più; anche Gemini 3 è decisamente più intraprendente di Gemini 2.5, essendo stato sviluppato per il programma di IA agentica Antigravity.



07

「Conclusioni」



In queste pagine abbiamo tracciato una panoramica sui sistemi di intelligenza artificiale e sul loro utilizzo in ambito aziendale. Che si parli dell'uso individuale degli LLM più noti o dell'implementazione di sistemi interni dell'azienda, avete trovato alcuni spunti utili per iniziare a orientarvi in questo mondo. Per quanto le informazioni qui contenute possano sembrare molte, la materia è vasta e ogni realtà ha esigenze diverse che richiedono approfondimenti specifici.

Questa guida è una bussola per il cambiamento ma per una trasformazione digitale davvero efficace serve il **supporto di un sistema strutturato** capace di accompagnare le imprese in tutte le fasi dell'innovazione.

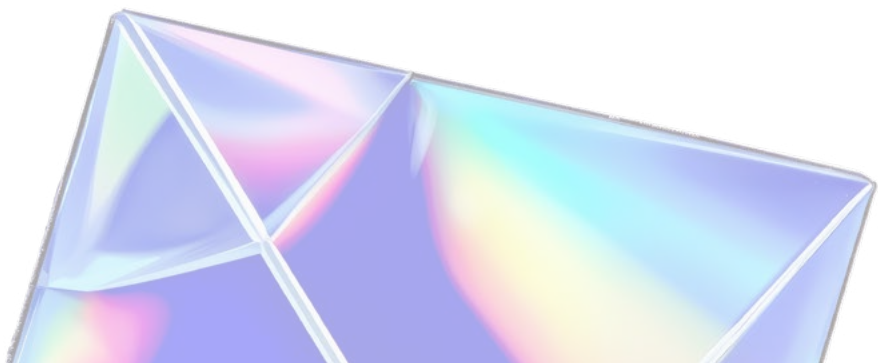
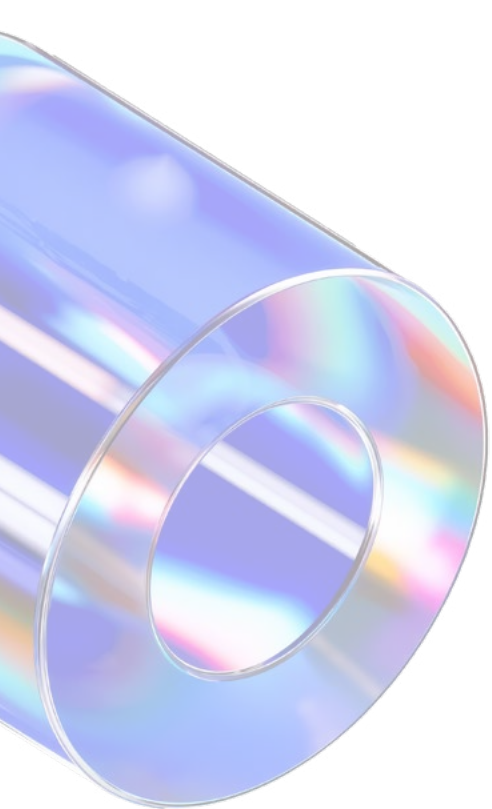
In questo contesto l'**Emilia-Romagna** rappresenta un ecosistema unico in Italia in termini di opportunità legate all'innovazione imprenditoriale, anche a vantaggio della **transizione digitale** delle imprese del territorio. Oggi, le aziende che desiderano avviare o consolidare percorsi di digitalizzazione possono contare sul supporto delle istituzioni regionali, che mette a disposizione strumenti, opportunità e competenze diffuse.

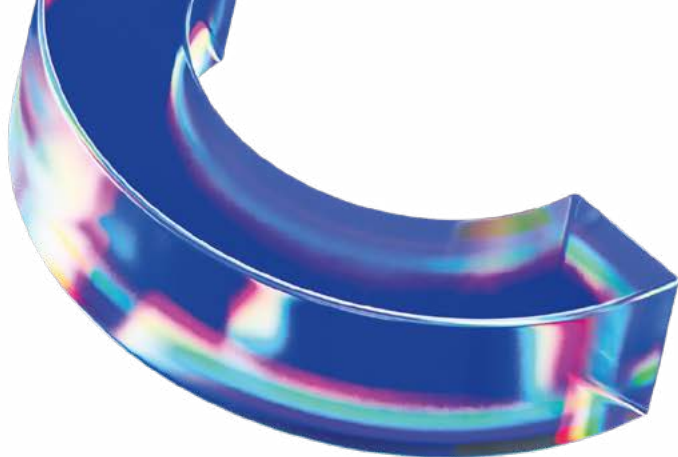
Citiamo, a titolo di esempio, la rete dei **Tecnopoli Emilia-Romagna** e dei **Laboratori della Rete Alta Tecnologia**, che offrono competenze specialistiche, strumentazioni avanzate, orientamento e servizi di ricerca industriale a beneficio delle imprese.

Parallelamente, la Regione promuove **voucher tecnologici e bandi** specificatamente dedicati alla transizione digitale, a sostegno di tecnologie 4.0, I.A. IoT, cloud, Big Data, automazione intelligente e cybersecurity. Sono **strumenti concreti** che incentivano l'innovazione e permettono anche alle imprese di compiere passi significativi.

Tutto questo si inserisce nel contesto della **Data Valley dell'Emilia-Romagna**, uno dei poli europei più importanti per il supercalcolo e la gestione dei dati. Il suo cuore è il **Tecnopolo DAMA**, che ospita al suo interno il supercomputer Leonardo e il centro meteo europeo ECMWF, creando un ambiente in cui imprese, università e centri di ricerca internazionali possono **collaborare** e sperimentare nuove soluzioni.

Grazie a questo insieme di investimenti, opportunità e infrastrutture, le imprese del territorio possono contare su un sistema di supporto **stabile e continuo**, in grado di seguirle in ogni fase del percorso di innovazione digitale dalla sperimentazione iniziale fino alla piena **trasformazione dei processi** e dei modelli di business. Un ecosistema che rende la transizione digitale non solo possibile ma concretamente accessibile.





Questa guida è stata realizzata da MISTER Smart Innovation s.c.r.l., nell'ambito del progetto G.A.I.A. - Generative AI per Imprese e Assessment, finanziato dal Bando per azioni di sistema a favore della rete regionale per la transizione digitale delle imprese - Edizione 2025 - PR FESR 2021-2027

Contenuti a cura di Carlo Foschi, con il contributo di Simone Lossano.

Redazione: Diana Tartaglia.

Grafiche e impaginazione: Valentina Mondini e Laura Bonazza.

