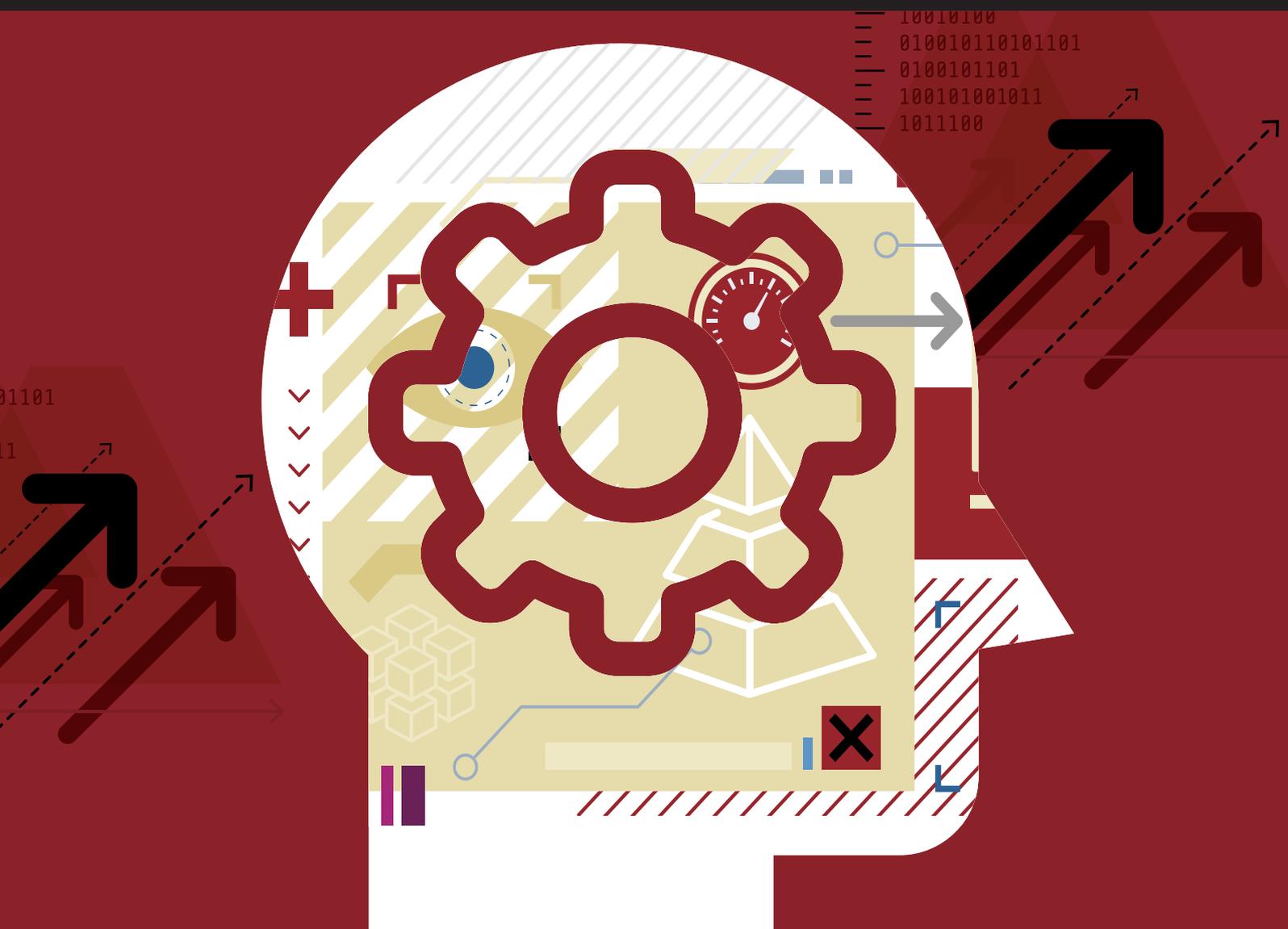


La grande accelerazione: le opinioni dei CIO sull'AI generativa



Premessa

“La grande accelerazione: le opinioni dei CIO sull'AI generativa” è un rapporto di MIT Technology Review Insights sponsorizzato da Databricks. Questo studio, basato su sette interviste approfondite con alti dirigenti ed esperti del settore, effettuate nei mesi di aprile e maggio 2023, ha l'obiettivo di capire come i leader tecnologici stiano adottando strumenti AI generativi nell'ambito della strategia di adozione dell'AI in tutta l'impresa.

Il rapporto attinge anche ai risultati di un sondaggio globale di MIT Technology Review Insights su 600 dirigenti nell'ambito della gestione dei dati e delle tecnologie) condotto nei mesi di maggio e giugno 2022. Il rapporto

è stato realizzato da Adam Green (autore), Teresa Elsey (revisore) e Nicola Crepaldi (editore). Lo studio è indipendente dal punto di vista editoriale e le opinioni in esso contenute sono attribuibili a MIT Technology Review Insights.

Desideriamo ringraziare le persone che hanno messo a disposizione il loro tempo e le loro informazioni:

Andrew Blyton, Vice President e Chief Information Officer, DuPont Water & Protection

Michael Carbin, Associate Professor, MIT, e Founding Advisor, MosaicML

Owen O'Connell, Senior Vice President and Chief Information Officer (Information Digital Services and Operations), Shell

Noriko Rzonca, Chief Digital Officer, Cosmo Energy Holdings

Richard Spencer Schaefer, Chief Health Informatics Officer, Kansas City VA Medical Center, e AI Solutions Architect, National Artificial Intelligence Institute, US Department of Veterans Affairs

Cynthia Stoddard, Senior Vice President e Chief Information Officer, Adobe

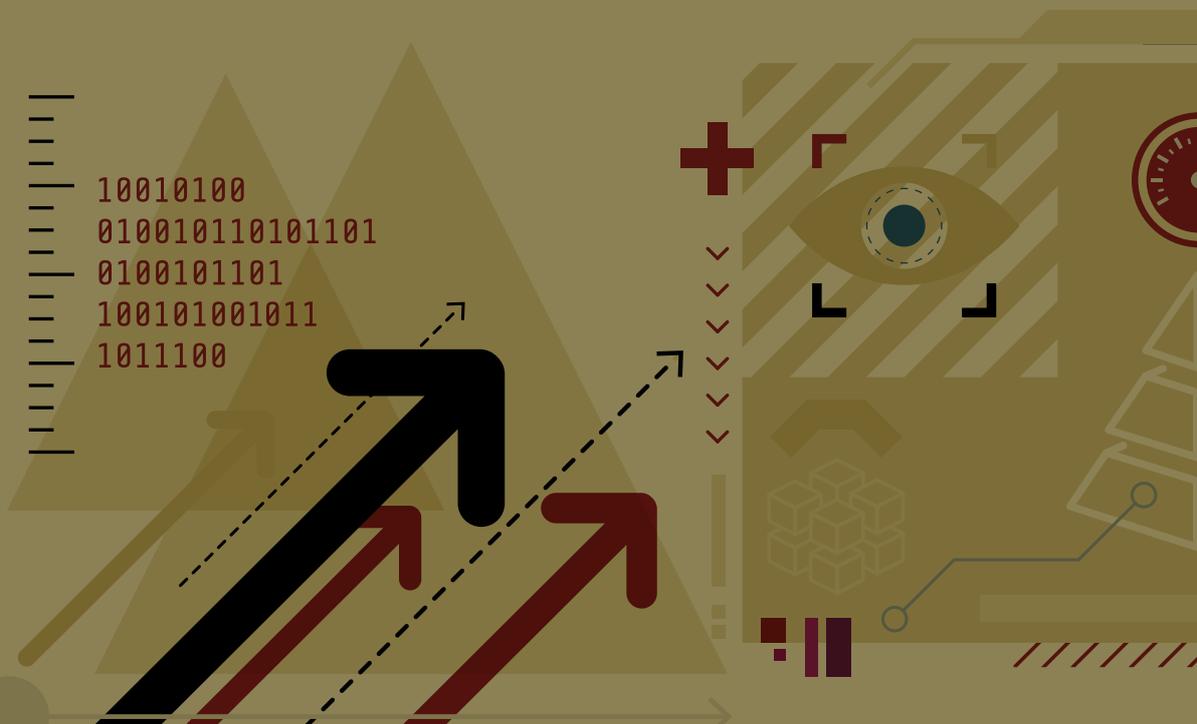
Matei Zaharia, Cofondatore e Chief Technology Officer (CTO), Databricks, e Associate Professor of Computer Science, University of California, Berkeley



10010100
010010110101101
0100101101
100101001011
1011100

SOMMARIO

01 Sintesi introduttiva	4
02 L'AI ovunque, subito	6
03 Costruire per l'AI	10
04 Acquistare o costruire? Aperto o chiuso?	13
05 Preoccupazioni per la forza lavoro?	16
06 Rischi e responsabilità	19
07 Conclusioni	22



01

Sintesi
introduttiva

L'avvento degli strumenti di AI generativa rivolti al consumatore fra la fine del 2022 e l'inizio del 2023 ha impresso una svolta repentina al dibattito pubblico sulla potenza e sul potenziale dell'intelligenza artificiale. L'AI generativa si era fatta largo fra gli esperti già con l'introduzione di GPT-2 nel 2019, ma solo ora il suo potenziale rivoluzionario è diventato chiaro alle imprese. L'impatto di questo evento (e l'effetto a cascata che innescherà) si farà sentire per molti decenni.

L'impatto dell'AI generativa su economie e imprese sarà rivoluzionario. Secondo le stime di McKinsey Global Institute, l'AI generativa porterà un valore aggiunto all'economia mondiale compreso fra 2.600 e 4.400 miliardi di dollari all'anno, aumentando complessivamente l'impatto economico dell'AI del 15-40%¹. La società di consulenza

prevede che l'AI automatizzerà metà di tutto il lavoro fra il 2040 e il 2060, sottolineando come l'AI generativa accorcerà questa finestra temporale di un decennio rispetto alle stime precedenti. Goldman Sachs prevede un aumento del 7% (o quasi 7.000 miliardi di dollari) del PIL mondiale attribuibile all'AI generativa e si aspetta che due terzi dell'occupazione negli Stati Uniti saranno interessati dall'automazione basata su AI².

I sistemi AI che generano testi, come il popolare ChatGPT, sono basati su modelli linguistici di grandi dimensioni (Large Language Models). I modelli LLM vengono addestrati su un corpus di dati enorme per rispondere a domande o eseguire compiti basati sulla probabilità statistica. Invece di cercare e sintetizzare risposte, utilizzano modelli matematici per prevedere la parola o l'azione successiva più probabile³.

“Non mi viene in mente nulla di più potente dall'invenzione del personal computer.”

Michael Carbin, Associate Professor, MIT, e Founding Advisor, MosaicML



“Quello che mi ha entusiasmato quando ho interagito per la prima volta con ChatGPT è stata la sua capacità di fare conversazione”, afferma Michael Carbin, professore associato al MIT e consulente fondatore di MosaicML. “Per la prima volta mi sono sentito come se potessi comunicare con un computer e lui potesse capire ciò che intendevo. Finalmente possiamo trasformare il linguaggio in qualcosa di comprensibile a una macchina. Non mi viene in mente nulla di più potente dall’invenzione del personal computer”.

Sebbene l'importanza strategica dell'AI sia stata riconosciuta prima che l'AI generativa salisse alla ribalta, il nostro sondaggio del 2022 indicava che i CIO avevano ambizioni limitate: mentre il 94% delle organizzazioni usava l'AI in qualche forma, solo il 14% puntava a diffondere l'AI a tutti i livelli dell'impresa entro il 2025. Per contro, la potenza degli strumenti di AI generativa per democratizzare l'AI (cioè diffonderla a ogni funzione aziendale per supportare ogni dipendente e coinvolgere ogni cliente) fa presagire il raggiungimento di un punto di svolta dove l'AI possa crescere da tecnologia impiegata per casi d'uso particolari a tecnologia che definisce realmente l'impresa moderna.

Chief Information Officer e direttori tecnici dovranno quindi agire in modo deciso: adottare l'AI generativa per coglierne le opportunità ed evitare di cedere terreno alla concorrenza, prendendo al tempo stesso decisioni strategiche su infrastruttura dati, proprietà del modello, struttura della forza lavoro e governance dell'AI che abbiano effetti a lungo termine sul successo dell'organizzazione.

Questo studio analizza le opinioni più recenti dei CIO di alcune delle aziende più grandi e famose del mondo, oltre che di esperti del settore pubblico e privato e del mondo accademico. Le loro idee sull'AI vengono illustrate sullo sfondo del nostro sondaggio globale condotto su 600 alti dirigenti nell'ambito di dati e tecnologie⁴.

Le principali evidenze dello studio sono le seguenti:

- **AI generativa e LLM stanno democratizzando l'accesso all'intelligenza artificiale, permettendo finalmente all'AI di iniziare a diffondersi in tutta l'azienda.** Alimentata dal potenziale di nuovi casi d'uso emergenti, l'AI si sta finalmente evolvendo da progetti pilota e “isole di eccellenza” a funzionalità generalizzata intrecciata nel tessuto dei flussi di lavoro di un'organizzazione. I team di tecnologia non devono più “vendere” l'AI alle business unit: esiste ormai una “domanda spontanea” significativa da parte dell'azienda.

- **Una miniera di dati non strutturati e sepolti è ora accessibile, generando valore per l'azienda.** Le precedenti iniziative AI si dovevano concentrare su casi d'uso nei quali i dati strutturati erano pronti e abbondanti; la complessità nel raccogliere, annotare e sintetizzare set di dati eterogenei rendeva impraticabili iniziative AI su più larga scala. Al contrario, la nuova capacità dell'AI generativa di portare in superficie e utilizzare dati finora nascosti favorirà straordinari progressi in tutta l'organizzazione.

- **L'era dell'AI generativa richiede un'infrastruttura dati che sia flessibile, scalabile ed efficiente.** Per sostenere queste nuove iniziative, CIO e responsabili tecnici stanno adottando infrastrutture di nuova generazione. Approcci più avanzati come il lakehouse possono democratizzare l'accesso a dati e analisi, aumentare la sicurezza e unire storage a basso costo con query ad alte prestazioni.

- **Alcune organizzazioni puntano a sfruttare tecnologie open-source per costruire i propri LLM e per capitalizzare e proteggere i propri dati e IP (proprietà intellettuale).** I CIO conoscono già i limiti e i rischi dei servizi di terze parti, incluso il rilascio di dati sensibili e la necessità di affidarsi a piattaforme che non possono controllare o su cui non hanno visibilità. Vedono inoltre opportunità legate allo sviluppo di LLM personalizzati e alla generazione di valore da modelli più piccoli. Le organizzazioni di maggior successo troveranno il giusto equilibrio strategico sulla base di un attento calcolo del rischio, del vantaggio competitivo e della governance.

- **L'allarme legato ai rischi dell'automazione non deve essere ignorato, ma le previsioni distopiche sono esagerate.** Gli strumenti di AI generativa possono già svolgere compiti complessi e variegati, ma i CIO e gli accademici intervistati per questo studio non temono minacce su larga scala legate all'automazione. Ritengono piuttosto che gran parte della forza lavoro verrà liberata da mansioni che implicano grande dispendio di tempo, per concentrarsi su attività di approfondimento, strategia e valore commerciale.

- **Una governance unificata e coerente rappresenta i binari lungo i quali l'AI può avanzare velocemente.** L'AI generativa comporta rischi commerciali e sociali, dalla protezione di IP sensibili alla violazione del diritto d'autore, da risultati inaffidabili o inspiegabili a contenuti tossici. Per innovare velocemente senza provocare danni o anticipare i cambiamenti normativi, i CIO più all'avanguardia devono affrontare le problematiche di governance dell'AI generativa, investendo in tecnologie, processi e strutture istituzionali.

02

L'AI ovunque, subito

Grazie alla sua flessibilità e portata, nonché all'interfaccia semplice basata sul linguaggio naturale, l'AI generativa sta dimostrando la sua solidità in ogni ambito, dalla scrittura di testi alla scrittura di codice. La sua potenza e il suo potenziale di rivoluzionare il modo in cui il lavoro viene svolto in ogni settore e funzione aziendale suggeriscono che l'AI generativa avrà un impatto simile a quello del personal computer, di Internet o dello smartphone, creando modelli di business completamente nuovi, dando vita a nuovi leader di settore e facendo dimenticare a tutti come si lavorava prima del suo avvento (vedi Figure 1 e 2).

Questo interesse improvviso per la potenza e il potenziale dell'AI ha segnato un punto di svolta nel modo in cui le imprese considerano l'AI: fino a poco tempo fa, l'adozione dell'AI era

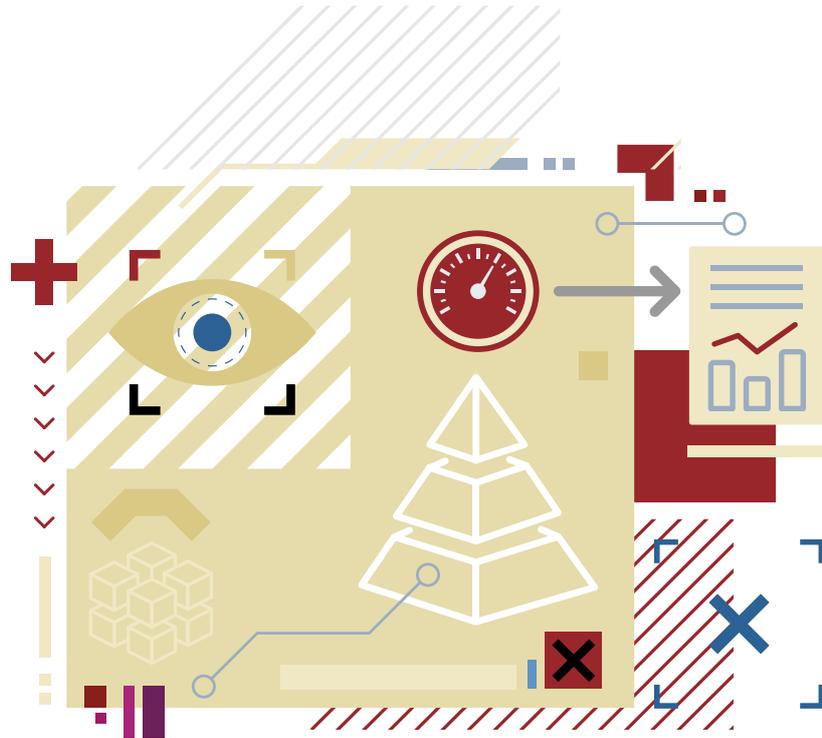
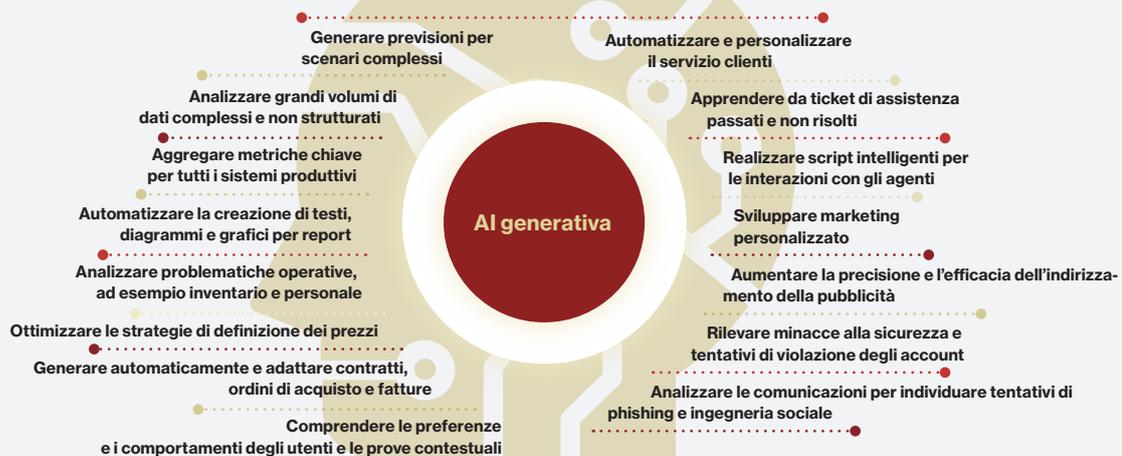


Figura 1: Applicazioni e casi d'uso dell'AI generativa nelle imprese

Esistono solo poche funzioni aziendali che fanno un uso significativo dell'AI generativa.



Fonte: Redatto da MIT Technology Review Insights sulla base dei dati pubblicati in "Retail in the Age of Generative AI"⁵, "The Great Unlock: Large Language Models in Manufacturing"⁶, "Generative AI Is Everything Everywhere, All at Once"⁷ e "Large Language Models in Media & Entertainment"⁸ Databricks, aprile-giugno 2023.

molto disomogenea fra i vari settori e fra le diverse funzioni all'interno di un'azienda. Nel nostro studio del 2022, solo l'8% degli intervistati ha dichiarato che l'AI era un elemento critico per tre o più funzioni aziendali. Solo nel settore finanziario e nell'IT più della metà degli intervistati ha risposto che l'AI era una tecnologia chiave o diffusa all'interno della funzione.

Prima dell'avvento dell'AI generativa, solo poche organizzazioni avevano adottato l'AI come funzionalità chiave in tutta l'azienda, o potevano aspirare a farlo. Mentre il 94% delle organizzazioni utilizzava già l'AI in qualche modo, solo il 14% puntava a realizzare l'AI "estesa a tutta l'impresa", che significa introdurre l'AI in almeno cinque funzioni chiave, entro il 2025 (vedi Figure 3 e 4).

Ora, però, l'AI generativa sta cambiando i termini del dibattito. Dimostrando di poter trovare applicazione in ogni funzione aziendale, l'AI è pronta a espandersi in tutta l'impresa. L'AI si sta espandendo nelle applicazioni creative, una volta considerate

dominio esclusivo dell'essere umano. Adobe, ad esempio, ha lanciato Firefly, una famiglia di modelli creativi di AI generativa che fungono da "copilota" per flussi di lavoro creativi e di grafica/design, come spiega Cynthia Stoddard, Senior Vice President e Chief Information Officer dell'azienda. Gli strumenti di Firefly possono ricolorare immagini esistenti, generare nuove immagini e aggiungere/togliere oggetti da immagini, tutto partendo da una descrizione testuale.

I settori energetico e chimico stanno applicando l'AI in ambiti finora inaccessibili. Il colosso multisettoriale Dupont, ad esempio, aveva già lavorato in passato su chatbot per dipendenti e consumatori, ma era rimasto deluso dalla loro imprecisione. "Ora gli LLM riescono a garantire la necessaria accuratezza, e in tempi più rapidi", sottolinea Andrew Blyton, Vice President and Chief Information Officer, DuPont Water & Protection. L'azienda sta utilizzando l'AI per applicazioni di programmazione della produzione, affidabilità e manutenzione predittiva, e ottimizzazione dei prezzi di vendita.

Figura 2: Applicazioni e casi d'uso per settori industriali specifici

Ogni settore (e ogni impresa) può trovare applicazioni per l'AI generativa.



Largo consumo e commercio al dettaglio

- Offrire camerini virtuali
- Pianificare consegne e installazioni
- Fornire assistenza per la ricerca di prodotti nel punto vendita
- Ottimizzare la previsione della domanda e la pianificazione dell'inventario
- Generare progetti di nuovi prodotti



Industria manifatturiera

- Fungere da "copilota esperto" per i tecnici
- Consentire conversazioni interattive con macchine
- Fornire assistenza sul campo prescrittiva e proattiva
- Risoluzione di problemi con linguaggio naturale
- Valutare lo stato e la documentazione di garanzia
- Individuare colli di bottiglia nei processi e definire strategie correttive



Media e intrattenimento

- Fornire ricerche intelligenti e scoperta di contenuti su misura
- Scrivere titoli e testi accattivanti
- Dare feedback in tempo reale sulla qualità dei contenuti
- Curare playlist personalizzate, notiziari e raccomandazioni/consigli
- Sviluppare trame interattive (storytelling) sulla base delle scelte degli spettatori
- Proporre offerte e piani di abbonamenti mirati



Servizi finanziari

- Individuare potenziali segnali di trading e allertare i trader che hanno posizioni vulnerabili
- Accelerare le decisioni di sottoscrizione
- Ottimizzare e ricostruire sistemi esistenti
- Sottoporre a ingegneria inversa modelli bancari e assicurativi
- Monitorare potenziali crimini e frodi finanziarie
- Automatizzare la raccolta dati per la conformità normativa
- Estrarre informazioni da dichiarazioni ufficiali dell'azienda

Figura 3: Il grande ripensamento

Prima dell'AI generativa, poche organizzazioni avevano adottato l'AI come parte fondamentale di una qualsiasi funzione aziendale.

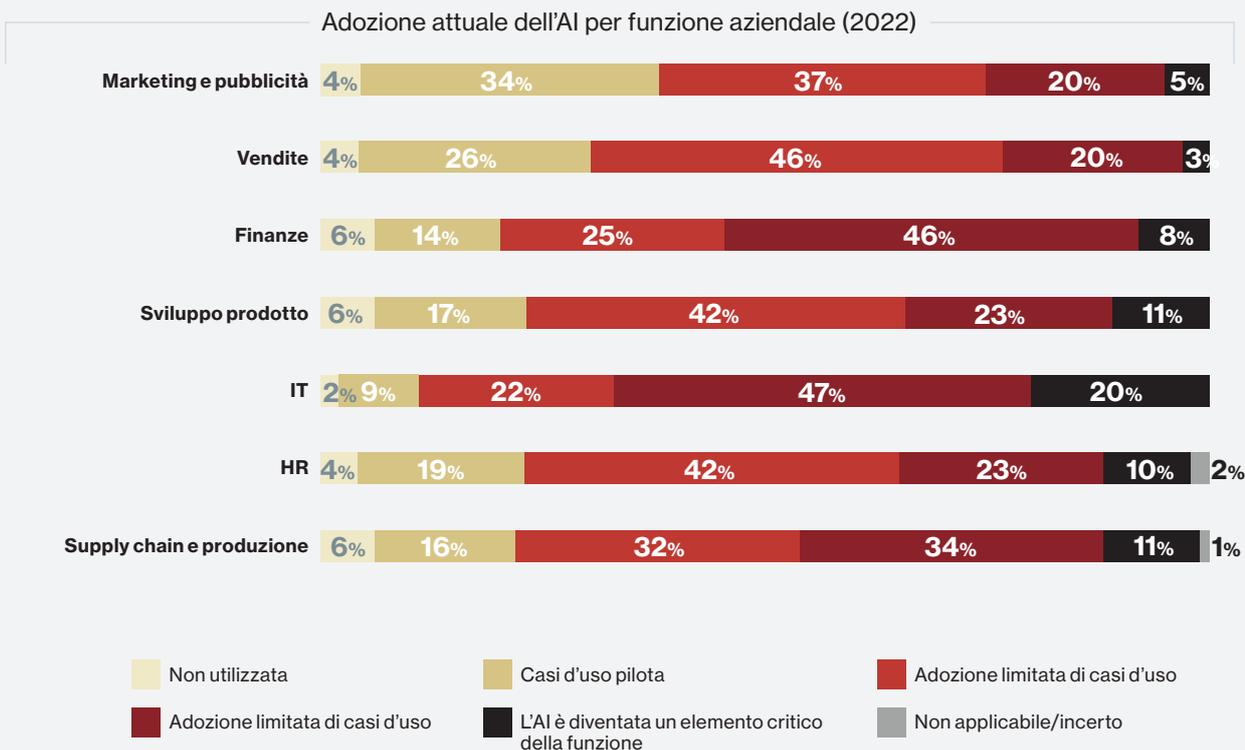
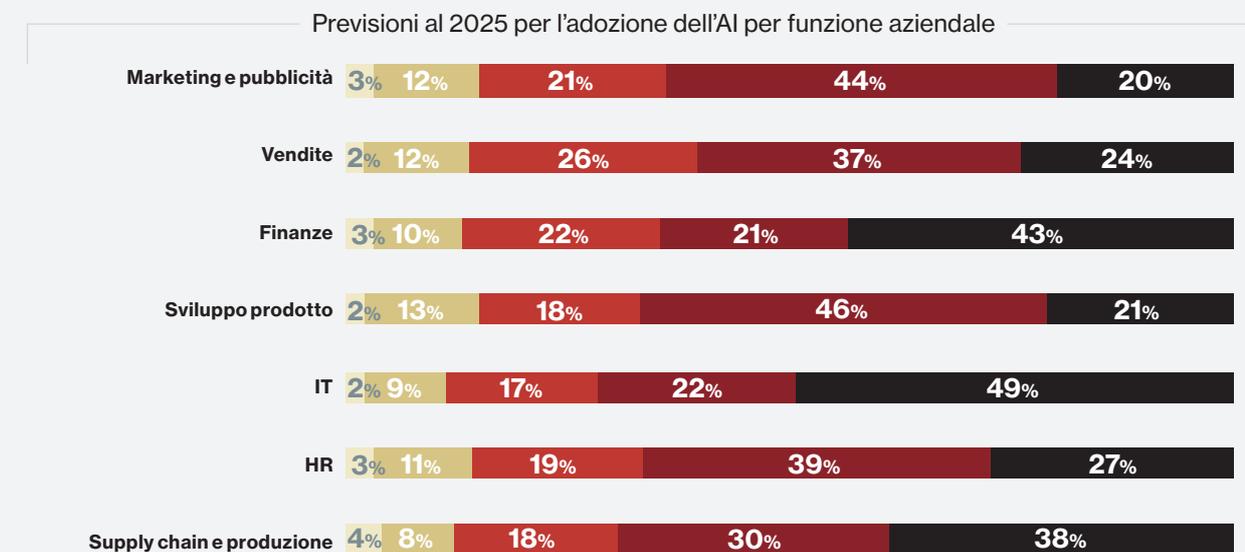


Figura 4: Ambizioni limitate per l'AI

Prima dell'AI generativa, poche organizzazioni puntavano ad adottare l'AI come funzionalità essenziale in reparti aziendali chiave entro il 2025.



Le multinazionali con tecnologie implementate nell'arco di diversi decenni hanno sempre faticato a unificare la loro infrastruttura digitale. Fusioni e acquisizioni hanno prodotto infrastrutture IT frammentarie. Documenti importanti, da studi di ricerca e sviluppo a istruzioni per la progettazione di impianti, sono rimasti nascosti, imprigionati in formati di file proprietari conservati offline. “Potremmo interrogare questi documenti usando modelli LLM? Possiamo addestrare modelli per fornirci informazioni che noi non riusciamo a vedere in questo mare di documentazione?” si chiede Blyton. “Pensiamo che sia un caso d'uso ovvio”. I modelli linguistici promettono di rendere questi dati non strutturati molto più preziosi.

Il colosso dell'energia Shell concorda: “Abbiamo cominciato a vedere i benefici perché molta della documentazione finora conservata negli archivi sta cominciando a emergere”, racconta Owen O'Connell, Senior Vice President and Chief

“Ora gli LLM riescono a garantire la necessaria accuratezza, e in tempi più rapidi”.

Andrew Blyton, Vice President e Chief Information Officer, DuPont Water & Protection

Information Officer per i servizi e le attività digitali. L'azienda sta anche snellendo tutta la documentazione cartacea relativa a leggi, normative e risorse umane per le sue numerose giurisdizioni, e persino estrapolando informazioni dettagliate e approfondite da dati non strutturati in ambiti quali assunzioni e prestazioni.

Grandi speranze per il settore sanitario

Anche la salute e la medicina, ambiti fortemente incentrati sull'essere umano, hanno offerto terreno fertile per la sperimentazione dell'AI. In laboratorio, gli strumenti supportati da AI hanno dimostrato la loro capacità di prevedere le strutture delle proteine, contribuire alla scoperta di farmaci e tracciare la diffusione di epidemie, incluso il Covid-19. Ora l'AI generativa sta dimostrando di poter assistere efficacemente il personale medico in prima linea. Gli strumenti di elaborazione del linguaggio naturale, ad esempio, possono trascrivere e riassumere note mediche, mentre i chatbot possono essere addestrati per rispondere a domande mediche dei consumatori.

Richard Spencer Schaefer, Chief Health Informatics Officer, Kansas City VA Medical Center, l'ente che gestisce la sanità per i veterani di guerra americani, e architetto di soluzioni AI presso il VA National Artificial Intelligence Institute, è ottimista sul contributo che l'automazione e l'analisi predittiva potrebbero dare al settore. Recentemente

il dipartimento VA ha validato un modello che identificava il rischio entro le 24 ore per un paziente ricoverato in ospedale dopo essere stato in terapia intensiva. “Abbiamo notato un notevole miglioramento nella precisione e nella prevedibilità di utilizzo di quel modello AI, al punto tale che potremmo avere una riduzione significativa del tasso di mortalità”, riferisce.

Schaefer è entusiasta anche della capacità dell'AI di alleviare la fatica del personale sanitario e, al tempo stesso, aumentare la loro fiducia grazie a piccoli miglioramenti operativi concreti. “Penso che l'ambito in cui assisteremo a un'adozione molto più rapida nel settore sanitario”, afferma, “saranno miglioramenti molto specifici nell'automazione di alcuni processi.” Il VA, ad esempio, sta lavorando a un progetto basato su AI per ridurre gli allarmi e gli avvisi inutili nei centri medici. I modelli di machine learning possono aiutare a definire soglie di intervento dinamiche e specifiche per paziente, invece di fare riferimento alle soglie e ai livelli statici dei sistemi di allarme tradizionali. “A questo scopo”, sottolinea Schaefer, “dobbiamo costruire la fiducia e aiutare il personale sanitario a capire quale possa essere l'utilità dell'AI”.

03 Costruire per l'AI

L'infrastruttura e l'architettura dati poggiano su una solida infrastruttura dati che consente la raccolta, lo stoccaggio e l'analisi di un vasto "dataverso". Una piattaforma unificata per l'analisi dei dati e l'AI è considerata cruciale da quasi il 70% del campione del nostro sondaggio (vedi Figura 5), prima ancora dell'emergere delle applicazioni aziendali dell'AI generativa nel 2022.

L'infrastruttura e l'architettura dati comprendono tutta l'infrastruttura relativa a software e rete, in particolare cloud o cloud ibrido, oltre all'hardware come GPU ad alte prestazioni. Le imprese hanno bisogno di un'infrastruttura che massimizzi il valore dei dati senza compromettere la protezione e la sicurezza, soprattutto in un momento in cui il manuale delle regole per la protezione dei dati e l'AI diventa sempre più voluminoso. Per democratizzare veramente l'AI, l'infrastruttura deve supportare un'interfaccia semplice che consenta di interrogare i dati ed eseguire compiti complessi utilizzando il linguaggio naturale. "L'architettura si sta evolvendo in un modo che favorisce la democratizzazione dell'analisi", afferma Schaefer.

“Abbiamo aggregato dati su molte tecnologie diverse nel tempo, e penso che oggi stiamo scoprendo che il lakehouse offre il miglior rapporto costi-prestazioni in assoluto”.

Andrew Blyton, Vice President e Chief Information Officer, DuPont Water & Protection

I data lakehouse sono diventati una scelta frequente per l'infrastruttura. Si tratta di una soluzione ibrida fra due approcci tradizionali: data warehouse e data lake. I data warehouse sono saliti alla ribalta negli anni Ottanta per sistematizzare le conoscenze e la reportistica delle imprese. Tuttavia, i data warehouse non offrono servizi in tempo reale, perché eseguono elaborazioni in batch, e non possono gestire formati di dati emergenti o non tradizionali. I data lake, preferiti per la loro capacità di supportare più compiti di AI e data science, sono arrivati in tempi più recenti. Ma sono complessi da costruire, lenti e penalizzati da un livello inferiore di controllo della qualità dei dati. Il lakehouse combina il meglio di entrambi, offrendo un'architettura aperta che unisce la flessibilità e la scalabilità dei data lake con la gestione e la qualità dei dati del data warehouse.



Controllare il “rubinetto” dei dati

Addestrare GPT-3 di Open AI è costato più di 4 milioni di dollari, mentre l'addestramento del modello LLaMA di Meta ha richiesto un milione di ore di GPU e oltre 2,4 milioni di dollari di costi¹³. L'addestramento di un modello AI consuma molta energia, con un fabbisogno di elettricità che può superare quello di 100 abitazioni americane in un anno¹⁴. L'addestramento di GPT-3 di Open AI ha consumato 1.287 MWh di elettricità ed emesso oltre 550 tonnellate di biossido di carbonio¹⁵. “Questi modelli sono fra i più pesanti che si possano addestrare”, afferma Owen O’Connell di Shell.

I modelli più grandi sono più costosi da addestrare ed eseguire, con un costo per ogni interazione sostanzialmente proporzionale alle dimensioni del modello, spiega Michael Carbin del MIT e di MosaicML. OpenAI ha dichiarato di spendere 40 milioni di dollari al mese per elaborare le domande degli utenti, mentre il chatbot Bing di Microsoft ha bisogno di un’infrastruttura da 4 miliardi di dollari per servire la sua base di utenti potenziale¹⁶. Modelli più piccoli e più efficienti (vedi il prossimo capitolo) possono rappresentare un’alternativa promettente.

Le aziende dovranno prestare maggiore attenzione all’efficienza, mentre startup e ricercatori stanno scoprendo nuovi modi di aumentare l’efficienza dei modelli AI. “Di fronte all’aumento dei costi, stiamo assumendo ingegneri che sappiano come ottimizzare e controllare i costi”, afferma Andrew Blyton di DuPont. “A un certo punto diventiamo molto più efficienti in termini di query o elaborazione perché possiamo ingaggiare esperti allo scopo di ottimizzare”. DuPont traccia e analizza i costi dell’elaborazione in cloud a livello granulare. “Abbiamo alcuni talenti interni molto dotati e ci affidiamo ai nostri fornitori per essere sicuri di ottimizzare ogni aspetto. In questo caso, un approccio sbagliato può avere un impatto significativo sia sui tempi, sia sulle prestazioni” aggiunge Blyton.

Per il VA, il lakehouse è interessante perché riduce al minimo lo spostamento dei dati, operazione che aumenta i rischi per la privacy e la governance. “Lo spostamento dei dati comporta notevoli problemi e, per questo motivo, impieghiamo molte risorse e competenze per gestire questo aspetto”, afferma Schaefer.

DuPont ha investito in molteplici infrastrutture tecnologiche e considera il data lakehouse molto promettente. “Abbiamo aggregato dati su molte tecnologie differenti nel tempo,” afferma Blyton” e penso che oggi stiamo scoprendo che il lakehouse offre il miglior rapporto costi-prestazioni in assoluto, pertanto abbiamo cominciato a investire più intensamente nella scalabilità”. Per DuPont, un data lakehouse offre visibilità su un portafoglio organizzativo complesso che comprende anni di attività di M&A e debito tecnico. “Dare un senso ai dati della nostra azienda è la ragione principale per cui investiamo nell’approccio del data lakehouse”, afferma Blyton. “Come possiamo prendere diversi sistemi gestionali (ERP) di dati, fonderli insieme e dare alle persone accesso in tempo quasi reale a informazioni che finora venivano gestite manualmente?”

Il lakehouse elimina la complessità con modalità che consentono agli utenti di eseguire attività avanzate indipendentemente dalla loro competenza tecnica. Shell ha costruito un “livello aziendale” che consente agli utenti di interagire dinamicamente. “In passato dovevamo accedere a data store, estrarre i dati, ripulirli ed eseguire svariate attività di trasformazione”, ricorda O’Connell. Il lakehouse consente agli utenti di eseguire carichi di lavoro, fare analisi storiche e comparare trend in totale autonomia, mentre i servizi in cloud forniscono l’elasticità necessaria per l’elaborazione.

Il data lakehouse supporta dati sia strutturati sia non strutturati, sottolinea O’Connell. “I database tradizionali sono progettati per un solo tipo di dati”, osserva. Il lakehouse “ci consente di agire molto più velocemente. Tutti i dati confluiscono in un’architettura comune e le persone li possono mescolare e ricavarne valore per l’azienda”. Si tratta di un’opportunità fondamentale per un’azienda ad alta intensità di dati come Shell. Per la sola manutenzione predittiva, l’azienda ha 17.000 modelli e 4.000 miliardi di righe di dati, a cui se ne aggiungono altri 20 miliardi al giorno dai 3 milioni di sensori installati negli impianti in tutto il mondo.

Chi adotta il data lakehouse utilizza questa tecnologia come parte integrante di una strategia dati che consente di realizzare casi d’uso di dati e AI pratici e di grande impatto. Dice Schaefer: “Democratizzare i dati è piuttosto costoso e

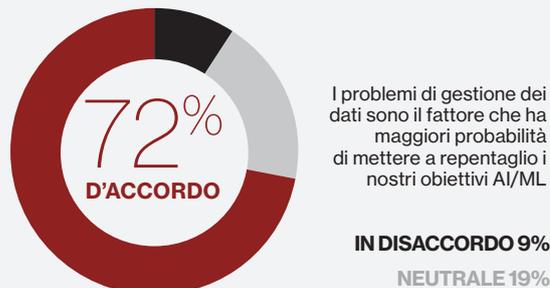
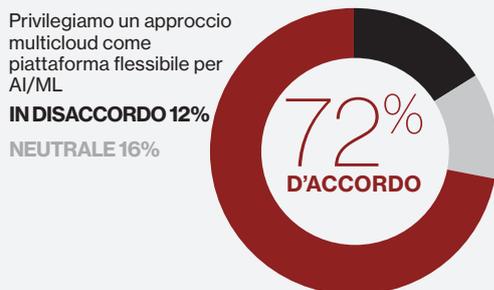
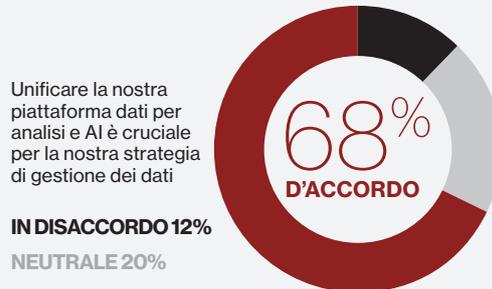
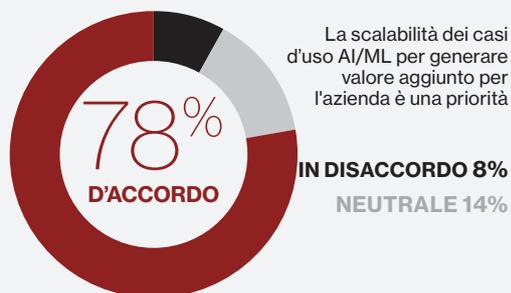
non ci sono mai state risorse efficaci a tale scopo. Stiamo però entrando rapidamente nell'era del lakehouse e di modelli transazionali e strumenti software ibridi che finalmente operano bene insieme. Penso che nella sanità in particolare assisteremo a un'adozione piuttosto rapida di strumenti che miglioreranno l'efficienza operativa, perché abbiamo molte opportunità di migliorare il modo in cui facciamo le cose e molte opportunità per l'automazione dei processi”.

Blyton aggiunge che, in DuPont, “stiamo utilizzando il lakehouse e strumenti come il data lake per costruire una piattaforma di dati e applicare su di essa algoritmi di ML. Stiamo potenziando la manutenzione predittiva utilizzando algoritmi con sensori di terze parti per capire quando una macchina sta per guastarsi, prima che si guasti e non dopo che si è guastata. Questi casi d'uso sono molto validi e concreti e il valore aggiunto generato dall'investimento produce un ritorno tangibile e misurabile per l'azienda”.

“I database tradizionali sono progettati per un solo tipo di dati. Il lakehouse ci consente di agire molto più velocemente.”

Owen O'Connell, Senior Vice President and Chief Information Officer (Information Digital Services and Operations), Shell

Figura 5: Priorità dell'AI per la dirigenza



04 Acquistare o costruire? Aperto o chiuso?

Oggi, i CIO e il Leadership Team stanno rivedendo le loro valutazioni su proprietà, partnership e controllo, pensando come sfruttare le funzionalità di piattaforme di AI generativa di terze parti.

Sovrasfruttare una piattaforma AI generica difficilmente porta un vantaggio competitivo. Secondo Carbin, “se ti sta molto a cuore un particolare problema o se intendi costruire un sistema davvero fondamentale per la tua attività, la questione è chi possiede la tua proprietà intellettuale”. DuPont è una “azienda di scienza e innovazione”, aggiunge Blyton, “e ha la necessità di mantenere i modelli LLM all’interno dell’organizzazione, per proteggere la nostra proprietà intellettuale... è un’esigenza fondamentale”.

Mettere a rischio la competitività è una delle preoccupazioni. “Non è necessario partire da un modello esistente nel quale i dati che vengono immessi potrebbero essere utilizzati dall’azienda titolare del modello per fare concorrenza ai tuoi prodotti principali”, afferma Carbin. Inoltre, gli utenti non hanno visibilità sui dati, sulle ponderazioni e sugli algoritmi alla base di modelli chiusi, e il prodotto e i suoi dati di addestramento possono cambiare in qualsiasi momento. Le preoccupazioni riguardano soprattutto l’attività di R&D scientifica, dove la riproducibilità è critica¹⁷.

Alcuni CIO stanno intervenendo per limitare l’utilizzo di piattaforme di AI generativa esterne. Samsung ha bandito ChatGPT dopo aver scoperto che i dipendenti lo utilizzavano per lavorare su codice sensibile. Alcune altre aziende, fra cui JP

“Se ti sta molto a cuore un particolare problema o se intendi costruire un sistema davvero fondamentale per la tua attività, la questione è chi possiede la tua proprietà intellettuale.”

Michael Carbin, Associate Professor, MIT,
e Founding Advisor, MosaicML



Morgan Chase, Amazon e Verizon, hanno introdotto limitazioni o divieti. “Non possiamo consentire che cose come la strategia aziendale di Shell passino per ChatGPT”, afferma O’Connell. Poiché i modelli LLM abbracciano la totalità dell’universo online, l’esperto ritiene che in futuro le aziende saranno più attente a ciò che mettono online: “Stanno cominciando a capire che... aspetta... qualcun altro farà molti soldi con i miei dati”.

Un’altra preoccupazione sono risultati imprecisi e inaffidabili. Gli LLM più grandi sono, in virtù delle loro stesse dimensioni, fuorviati da false informazioni pubblicate online. Questo aspetto rafforza la posizione di chi chiede approcci più mirati, secondo Matei Zaharia, Cofondatore e Chief Technology Officer (CTO), Databricks, e Associate Professor of Computer Science, University of California, Berkley. “Se ti muovi in un ambito più ristretto”, spiega, “puoi evitare tutta la spazzatura e le informazioni indesiderate che circolano sul web”.

Le aziende non possono semplicemente produrre versioni proprie di questi modelli grandissimi. Per dimensioni e costi, questo tipo di lavoro computazionale va oltre le capacità di tutte le organizzazioni, salvo le più grandi: OpenAI ha dichiarato di aver utilizzato 10.000 GPU per addestrare ChatGPT¹⁸. Al momento, costruire modelli su larga scala è un’impresa riservata esclusivamente alle aziende tecnologiche meglio attrezzate.

Modelli più piccoli, invece, rappresentano un’alternativa praticabile. “Credo che passeremo da “mi serve un mezzo migliaio di miliardi di parametri in un modello” a “forse mi servono 7, 10, 30, 50 miliardi di parametri per i dati che ho realmente”, osserva Carbin. “Questa riduzione della complessità si ottiene restringendo il raggio d’azione da un modello tuttofare che contiene tutto lo scibile umano a conoscenze di altissima qualità specifiche per ogni azienda, perché questo è ciò di cui hanno effettivamente bisogno le persone nelle imprese”.

Fortunatamente, più piccolo non significa più debole. I modelli generativi sono stati perfezionati per settori che richiedono meno dati, come dimostrano modelli quali BERT, che offre versioni per contenuti biomedicali (BioBERT), legali (LegalBERT) e testi in francese (battezzato col delizioso nome di CamemBERT)¹⁹. Per casi d’uso specifici, le organizzazioni possono scegliere di scambiare l’ampiezza delle conoscenze con una maggiore specializzazione nel loro settore di attività. “Le persone stanno realmente cercando modelli che siano “avvezzi” al loro settore”, afferma Carbin. “Una volta fatto questo passo, si comincia a capire che si può operare in modo diverso e avere successo”.

Dolly: LLM conversazionale da 30 dollari

A marzo 2023, Databricks ha rilasciato Dolly, un LLM open-source progettato per democratizzare la potenza dell’AI generativa²⁰. Addestrato con meno di 30 dollari, ha dimostrato capacità conversazionali simili a ChatGPT (cioè la capacità di seguire le istruzioni dell’utente). Dolly è basato sullo strumento LLaMA di Meta, perfezionato con input di alta qualità raccolti dai dipendenti di Databricks in modalità di crowdsourcing. Dolly, così chiamato in ricordo del primo mammifero clonato, ha solo 6 miliardi di parametri, cioè lo 0,0034% dei 175.000 miliardi utilizzati in GPT-3.

“I modelli di fascia più alta, con centinaia di miliardi di parametri, possono essere molto costosi da addestrare ed eseguire”, sottolinea Matei Zaharia di Databricks. Questi costi spingono molte aziende ad affidarsi a strumenti proprietari come ChatGPT o a spendere ingenti somme e risorse per costruire modelli propri. L’aumento dell’efficienza di hardware e software farà diminuire i costi nel tempo, ma la costruzione di modelli con migliaia di miliardi di parametri resterà fuori dalla portata di molti, osserva.

Per evitare che gli strumenti di AI finiscano per essere concentrati nelle mani di poche grandi aziende, Databricks ha cercato di sviluppare un modello di chat open-source che utilizza codice e dati open-source ma consente anche utilizzi commerciali. “L’aspetto principale che volevamo indagare”, prosegue Zaharia, “è quanto costa creare un modello utilizzando solo set di dati pubblici disponibili liberamente, utilizzando codice open-source e utilizzando o raccogliendo dati di addestramento che siano open-source e possano essere eseguiti da chiunque”.

Il codice di addestramento, il set di dati e i pesi del modello utilizzati da Databricks per Dolly provengono tutti dal mondo open-source. La release di aprile 2023, Dolly 2.0, è il primo LLM open-source concesso in licenza per uso commerciale²¹. Questo consente alle aziende di combinare i loro dati con il set di dati di Databricks, per costruire applicazioni personalizzate che non compromettano la loro IP o informazioni aziendali.

Consentendo alle aziende di costruire e personalizzare strumenti propri a basso costo, le piattaforme come Dolly possono democratizzare l’accesso all’AI generativa.

“Tutti i modelli di grandi dimensioni che si possono avere da terze parti sono addestrati su dati presenti nel web. Ma ogni organizzazione possiede molti concetti e dati che questi modelli non conoscono.”

Matei Zaharia, Cofondatore e Chief Technology Officer (CTO), Databricks, e Associate Professor of Computer Science, University of California, Berkley

“Le persone cominciano a considerare sempre di più i dati come un fossato economico”, spiega Zaharia. “Esempi come BloombergGPT [un LLM specifico per la finanza] indicano che le aziende stanno valutando cosa fare con i loro dati e stanno commercializzando i propri modelli”.

“Le aziende amplieranno e personalizzeranno questi modelli con i loro dati, anche allo scopo di integrarli nelle proprie applicazioni affinché abbiano senso per il loro business”, prevede Zaharia. “Tutti i modelli di grandi dimensioni che si possono avere da terze parti sono addestrati su dati presenti nel web. Ma ogni organizzazione possiede molti concetti e dati che questi modelli non conoscono. L’aspetto interessante è che il modello non ha bisogno di un’enorme quantità di dati o tempi di addestramento aggiuntivi per apprendere nozioni su un nuovo comparto”, sottolinea.

Modelli open-source più piccoli come LLaMA di Meta potrebbero competere con le prestazioni di modelli di grandi dimensioni e consentire ai professionisti di innovare, condividere e collaborare. Il nostro team ha costruito un LLM utilizzando i pesi di LLaMA e spendendo meno di 600 dollari, rispetto ai 100 milioni di dollari investiti nell’addestramento di GPT-4. Questo modello, chiamato Alpaca, ha prestazioni equivalenti all’originale per certi compiti²². Inoltre, grazie alla maggiore trasparenza dell’open source, ricercatori e utilizzatori possono individuare più facilmente pregiudizi e difetti di questi modelli.

“Una buona parte di questa tecnologia può essere alla portata di molte più organizzazioni”, afferma Carbin. “Non solo realtà come OpenAI, Google e Microsoft, ma aziende di medie dimensioni e persino startup”.



05 Preoccupazioni per la forza lavoro

Dal telefono al computer desktop, ogni progresso della tecnologia nella vita quotidiana scatena preoccupazioni per la disoccupazione e la minaccia alle capacità e alle competenze degli esseri umani. L'AI generativa sembra inaugurare una nuova fase, considerando il ventaglio senza precedenti di attività che possono essere automatizzate in modo fattibile. Oggi, alcuni dei settori più vulnerabili all'avvento dell'automazione rientrano nell'ambito delle alte tecnologie e della creatività.

Secondo uno studio di Accenture, il 40% delle ore lavorative in tutti i settori potrebbe essere automatizzato o "aumentato" dall'AI generativa, laddove il potenziale maggiore risiede nei settori bancario e assicurativo, nei mercati di capitali e nel software²³. McKinsey prevede che l'AI generativa e le tecnologie collegate potrebbero automatizzare attività che attualmente assorbono il 60-70% del tempo dei lavoratori, rispetto al 50% precedente all'avvento dell'AI generativa²⁴.

Goldman Sachs prevede che due terzi dei posti di lavoro negli Stati Uniti saranno interessati dall'avvento dell'automazione con AI, ma non si aspetta che questo porti a un massiccio aumento della disoccupazione: "La maggior parte dei lavori e dei settori", secondo la società, "è esposta all'automazione solo parzialmente, pertanto non verrà sostituita dall'AI ma solo integrata". Inoltre, la perdita di posti di lavoro causata dalla tecnologia tende a essere bilanciata (e superata) dall'aumento dei lavori associati alla tecnologia: "Oltre l'85% dei posti di lavoro creati negli ultimi 80 anni riguarda nuove posizioni nate dal progresso tecnologico", scrivono gli autori²⁵.

I CIO e gli accademici intervistati per questo studio hanno una visione ottimistica sugli effetti dell'AI generativa per le imprese e la società nel suo complesso. Ritengono infatti che l'AI potrebbe aiutare settori come la sanità dove la forza lavoro è scarsa, e che l'esperienza umana resterà fondamentale. "Le persone temono di perdere il lavoro a causa dell'AI, ma nella sanità non è questa la preoccupazione principale", afferma Schaefer. "Le tecnologie che stiamo introducendo consentono ai medici di partecipare attivamente allo sviluppo dell'AI e, poiché è necessaria la convalida da parte di persone competenti, penso che ci sarà maggior fiducia nei modelli che sviluppiamo". Gli esseri umani

“Le tecnologie che stiamo introducendo consentono ai medici di partecipare attivamente allo sviluppo dell'AI e, poiché è necessaria la convalida da parte di persone competenti, penso che ci sarà maggior fiducia nei modelli che sviluppiamo”.

Richard Spencer Schaefer,
Chief Health Informatics Officer,
Kansas City VA Medical Center

sono ancora necessari per annotare, curare, esporre i modelli ai set di dati e garantire un'AI responsabile. “Abbiamo la possibilità di arrivare a un punto in cui potremo sfruttare l'AI per fare le nostre ricerche, perché c'è davvero molta collaborazione al momento”, aggiunge O'Connell.

L'AI generativa darà un contributo fondamentale anche democratizzando l'accesso a funzionalità tecniche finora limitata a una piccola fetta della forza lavoro. Le aziende stanno già adottando misure per favorire il coinvolgimento. “Abbiamo creato community di professionisti della sanità che non hanno necessariamente una formazione IT”, spiega Blyton. “Vogliamo mettere i nostri utenti nelle condizioni di creare i loro dashboard ed estrapolare informazioni dai dati”. Questo tipo di democratizzazione dei dati era già una delle modalità principali con cui le aziende ottenevano benefici concreti dall'AI nel 2022, secondo il nostro sondaggio (vedi Figura 6).

Blyton prevede anche che nuove idee per l'AI cominceranno ad arrivare dalla forza lavoro stessa, segnando l'inizio di un'era più improntata al self-service e all'intraprendenza. Secondo O'Connell, i responsabili delle business unit, ormai consapevoli del ruolo dell'AI grazie alla massiccia copertura mediatica degli ultimi mesi, stanno già creando una “domanda proattiva” di applicazioni AI, invece di continuare a vendere all'organizzazione le idee del team che si occupa di tecnologie.

Nell'ultimo decennio, il programmatore software è stata una delle figure più ricercate e pagate. Tuttavia, questo è uno dei campi in cui gli strumenti AI generativi si stanno dimostrando più utili, causando non pochi timori. I CIO e i responsabili tecnici intervistati per il nostro studio offrono una visione più ottimistica. Prevedono infatti un aumento della domanda di figure specializzate in efficienza operativa e AI responsabile e credono che i programmatori sposteranno la loro attenzione su lavori meno noiosi e con maggior valore aggiunto. Blyton chiede: “Abbiamo veramente bisogno di qualcuno che immette codice per sei ore in un motore e poi cerca di correggerlo per tre giorni? Eliminando questo processo, avremo molta più efficienza”.

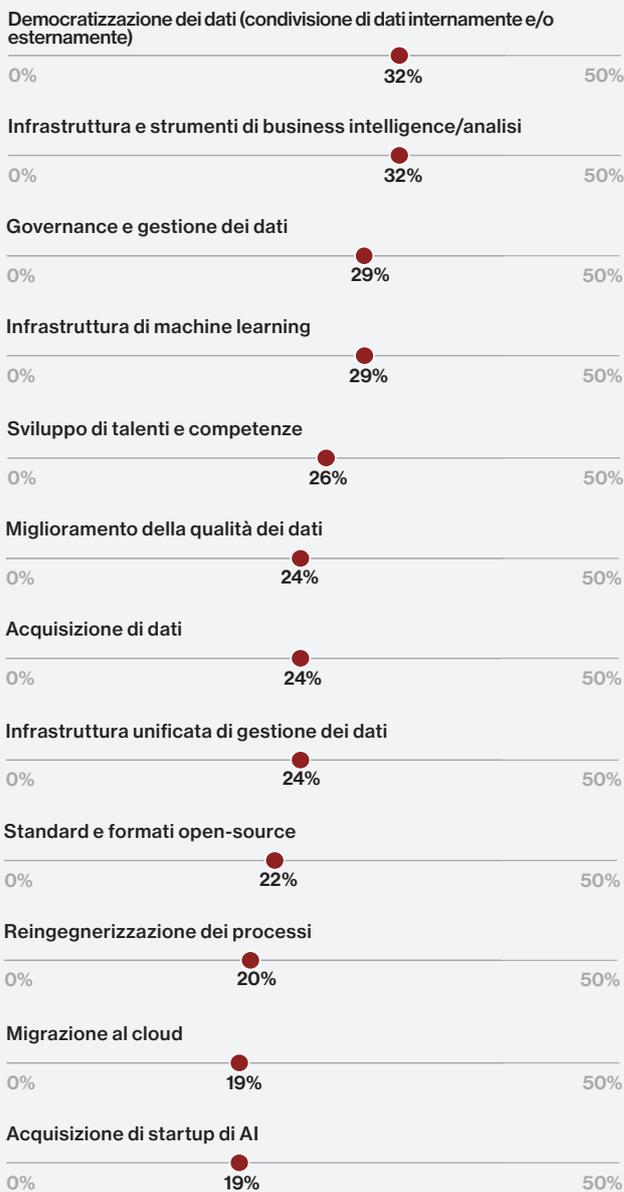
Alcuni si aspettano che i programmatori propongano idee grezze che vengano poi trasformate in oggetti più maturi e pronti all'uso. “Non è difficile immaginare che, in un futuro prossimo, potremmo utilizzare la lingua di tutti i giorni per generare codice”, prevede Blyton. “Entro un anno circa, il mio flusso di lavoro potrebbe prevedere utenti che svolgono parte dell'analisi in inglese, dopodiché io prelevo il codice in uscita e lo inserisco in qualcosa di più solido”. Secondo Stoddard l'AI sta accelerando una tendenza già in atto: “Servono persone

Figura 6:

Aree di investimento più utili per dati e AI

Alla domanda su quali siano stati gli investimenti più funzionali all'ottenimento di benefici da AI/ML, la risposta più frequente dei responsabili delle aziende è che la democratizzazione dell'accesso ai dati e la costruzione di un'infrastruttura di business intelligence hanno già ripagato l'investimento.

Percentuale di intervistati che classificano ciascuna opzione fra i loro tre investimenti più efficaci



Fonte: Sondaggio MIT Technology Review Insights, 2022.



“Nei prossimi cinque-dieci anni vedremo quanto sarà veloce l’adattamento, e le aziende che non si adatteranno sono destinate a scomparire, indipendentemente dalle dimensioni”.

Noriko Rzonca, Chief Digital Officer, Cosmo Energy Holdings

che conoscano la matematica e le scienze per lavorare sui modelli e altre che abbiano modalità di pensiero profondamente analitiche. Nel mio settore, i ruoli si sono evoluti verso l’analisi, il lavoro sui modelli, la valutazione degli scostamenti e la scrittura di codice”.

Gran parte dei CIO vede l’AI come un “copilota” piuttosto che un rivale degli addetti in carne e ossa. Zaharia prevede che i modelli LLM consentiranno ai dirigenti di acquisire dati di alta qualità per le loro decisioni, riducendo i tempi che gli analisti dovrebbero altrimenti impiegare per rispondere alle loro richieste di report specifici. Ma aggiunge: “Non penso che saranno in grado di svolgere in modo completamente automatico lavori estremamente complicati che richiedano molta pianificazione. Non penso che ci sia qualcuno che svolga un lavoro così semplice da poter essere interamente affidato a un modello linguistico”. Stoddard è d’accordo: “L’AI generativa consente a creatori di tutti i livelli di generare contenuti con le proprie parole”, spiega. “Ma non penso che rimpiazzerà mai gli esseri umani. Piuttosto, fornirà assistenza agli esseri umani”.

E aggiunge: “Internamente consideriamo AI/ML come un ausilio, un aiuto concreto ai nostri collaboratori, che consente loro di dedicare più tempo ad attività a valore aggiunto”. Adobe ha dovuto implementare un processo di gestione del cambiamento per conquistare la fiducia delle persone. Racconta Stoddard: “Quando abbiamo cominciato, le persone erano scettiche, ma quando hanno capito che potevano veramente dedicare più tempo ad attività a valore aggiunto e, in alcuni casi, le loro mansioni sono migliorate, tutti hanno creduto e continuano a credere nel progetto”. La democratizzazione dell’accesso a strumenti tecnici potrebbe coinvolgere tutta la società. “Per tutti coloro che non sono in

grado di programmare o manipolare software di design e grafica, quali opportunità ci sono?” chiede Carbin. “Questa tecnologia abbasserà la barriera di ingresso alla tecnologia e al software di programmazione o manipolazione. Oggi esiste un’interfaccia in linguaggio naturale per il software di grafica che consente di utilizzare uno strumento finora estremamente difficile da maneggiare, limitato ai migliori creativi professionisti”.

L’avversione al rischio e fattori culturali come la paura del fallimento devono essere ugualmente affrontati per promuovere l’adozione dell’AI da parte della forza lavoro. “A volte, chi ha lavorato o avuto ruoli dirigenziali in aziende giapponesi per lungo tempo, ed è quindi abituato a un certo stile, non è disposto ad assumersi rischi come gli addetti di più recente assunzione, che propongono continuamente nuove idee e sono aperti al cambiamento”, afferma Noriko Rzonca, Chief Digital Officer, Cosmo Energy Holdings.

“Un’azienda può aver bisogno di più tempo per adattarsi a nuove iniziative e metterle in atto, ma siamo fortunati sapendo che le aziende giapponesi sono molto brave a gestire il cambiamento. Man mano che ci adatteremo all’uso dell’AI in azienda, sono certo che vedremo grandi progressi e cambiamenti nel modus operandi delle imprese giapponesi. Credo che il Giappone abbia un enorme potenziale nell’ambito dell’intelligenza artificiale”.

Rzonca osserva che le organizzazioni dovranno abbracciare dati e AI se vogliono crescere: “Nei prossimi cinque-dieci anni vedremo quanto sarà veloce l’adattamento, e le aziende che non si adatteranno sono destinate a scomparire, indipendentemente dalle dimensioni”.

06

Rischi e responsabilità

L'AI, in particolare l'AI generativa, porta con sé problematiche di governance che potrebbero andare oltre le capacità delle infrastrutture esistenti per la governance dei dati. Quando si lavora con modelli generativi che “ingoiano e rigurgitano” tutti i dati immessi, senza tener conto della sensibilità dei dati stessi, le organizzazioni devono trovare nuove modalità per gestire il problema della sicurezza e della privacy. Le imprese devono anche gestire sorgenti di dati che crescono in modo esponenziale e dati generati da macchine o di provenienza discutibile, che richiedono quindi un approccio di governance unificato e coerente. Legislatori ed enti regolatori sono consapevoli dei rischi dell'AI generativa e, di conseguenza, hanno istruito casi legali e introdotto restrizioni e nuovi regolamenti, come ad esempio l'Artificial Intelligence Act dell'Unione Europea.

In questo contesto, i CIO sarebbero quantomeno incauti se adottassero strumenti AI senza gestirne i rischi, dai pregiudizi (bias) alla violazione del diritto d'autore, dalla riservatezza alle violazioni di sicurezza. In Shell “abbiamo dedicato molto tempo alle questioni legali e finanziarie e di riservatezza dei dati, etica e conformità, rivedendo le nostre politiche e infrastrutture per essere sicuri che siano pronte e adattate al nuovo contesto”, racconta O'Connell.

“Se il business model di un'azienda si basa interamente sulla sua proprietà intellettuale, la sicurezza è tutto.”

Andrew Blyton, Vice President e Chief Information Officer, DuPont Water & Protection

Una delle preoccupazioni riguarda la protezione della privacy nel momento in cui emergono e diventano utilizzabili quantità ingenti di nuovi dati. “Poiché la tecnologia è in una fase iniziale, c'è maggior bisogno di grandi set di dati per addestramento, convalida, verifica e analisi delle derive”, osserva Schaefer. Nel dipartimento VA, “questo scoperchia un vaso di Pandora per quanto riguarda la protezione dei dati sensibili dei pazienti. Abbiamo investito pesantemente in risorse cloud con funzionalità di governance federata e alta capacità di calcolo assicurata”.

La tutela della privacy commerciale e della proprietà intellettuale è una preoccupazione collegata. “Se il business model di un'azienda si basa interamente sulla sua proprietà intellettuale, la sicurezza è tutto”, afferma Blyton di DuPont. “Esistono molti malintenzionati che vogliono mettere le mani sulla nostra documentazione interna, pertanto l'apertura di nuovi varchi per trafugare la nostra IP è sempre fonte di preoccupazione”.

Un altro problema di governance riguarda l'affidabilità. I modelli LLM sono motori di apprendimento che raccolgono e sintetizzano contenuti da vasti archivi senza saper distinguere fra vero e falso. “Se le informazioni sono imprecise o obsolete, il modello comunque le memorizza. In un contesto aziendale, questo costituisce un problema e significa che le aziende devono prestare molta attenzione ai contenuti che finiscono nel modello”, avverte Zaharia. Blyton aggiunge: “ChatGPT è interessante in tal senso: se poni due volte la stessa domanda, ottieni risposte leggermente differenti, e questo fa storcere il naso a chi lavora in aziende di scienza e innovazione”.

La spiegabilità dei modelli è fondamentale per conquistare la fiducia delle figure coinvolte nell'adozione dell'AI e per dimostrare il valore della tecnologia per l'impresa. Questo tipo di spiegabilità è già oggi una priorità per la governance e la regolamentazione dell'AI, soprattutto dal momento in cui gli algoritmi hanno cominciato ad avere un ruolo in decisioni

importanti per la vita delle persone, ad esempio valutazioni della solvibilità creditizia o del rischio di reiterazione del reato. I critici sostengono che i sistemi di AI utilizzati in ambiti sensibili o di pubblico interesse non dovrebbero essere “scatole nere” e, di conseguenza, la revisione degli algoritmi è tema di crescente attenzione.

“È piuttosto evidente che la spiegabilità dei modelli nella sanità ha molte implicazioni”, osserva Schaefer. “Stiamo lavorando su schede di modelli”, (un tipo di documentazione di governance che fornisce informazioni standardizzate sull’addestramento, i punti di forza e le debolezze di un modello), “oltre che su strumenti per registri di modelli che offrano un certo livello di spiegabilità. Naturalmente la selezione degli algoritmi è un aspetto importante per la spiegabilità dei modelli”.

In Cosmo Energy Holdings, Rzonca sta valutando come bilanciare una solida governance dell’AI con le funzionalità offerte agli utenti. “Stiamo sviluppando regole di governance, le introduciamo e addestriamo le persone”, spiega. “Nel contempo, stiamo cercando di ottenere qualcosa a cui sia facile adattarsi, in modo che le persone possano percepire e vedere risultati concreti”. Per ottenere questo equilibrio, Schaefer mette in relazione l’accesso democratizzato ai dati e all’AI con la

“L’AI generativa aumenta esponenzialmente la possibilità e la promessa dell’AI. Si può trasformare la conversazione fra il creatore e il computer.”

Cynthia Stoddard, Senior Vice President e Chief Information Officer, Adobe

Il copilota creativo: l’AI in Adobe

I settori creativi stanno abbracciando l’AI generativa per offrire ai professionisti della grafica e del design nuove capacità affascinanti. La società di software Adobe propone una suite di modelli di grafica con AI generativa chiamata Adobe Firefly, che offrono capacità creative che proiettano verso nuovi confini le funzionalità di Photoshop, Illustrator, Express e tutte le soluzioni per le imprese.

L’azienda ha sviluppato anche Adobe Sensei GenAI per aumentare le velocità e la produttività in Adobe Experience Cloud, come copilota per i flussi di lavoro relativi alla customer experience. Questo consente di aumentare la produttività e l’efficienza in ambiti quali contenuti personalizzati, editing ed esperienza conversazionale.

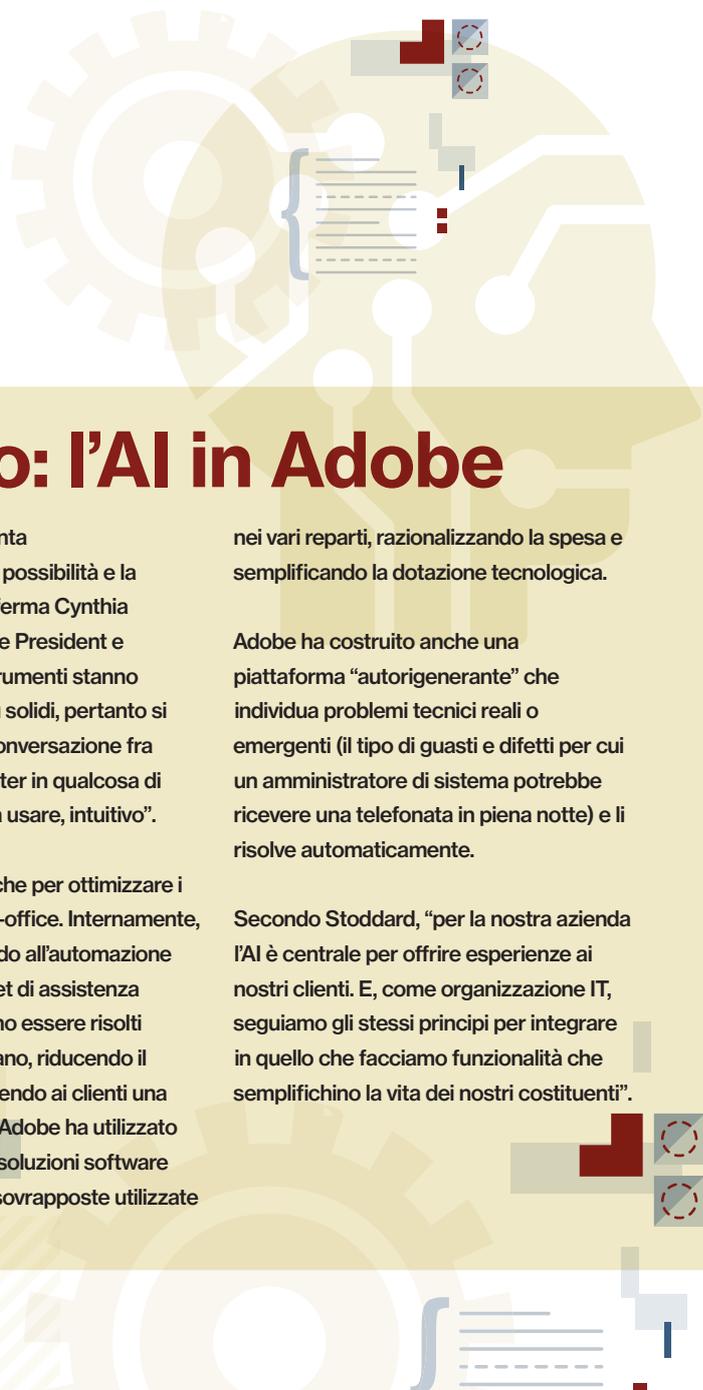
“L’AI generativa aumenta esponenzialmente la possibilità e la promessa dell’AI”, afferma Cynthia Stoddard, Senior Vice President e CIO di Adobe. “Gli strumenti stanno diventando molto più solidi, pertanto si può trasformare la conversazione fra il creatore e il computer in qualcosa di facile, più naturale da usare, intuitivo”.

Adobe utilizza l’AI anche per ottimizzare i suoi processi di back-office. Internamente, l’azienda sta ricorrendo all’automazione per selezionare i ticket di assistenza dei clienti che possono essere risolti senza intervento umano, riducendo il carico di lavoro e offrendo ai clienti una soluzione più rapida. Adobe ha utilizzato l’AI per catalogare le soluzioni software numerose e spesso sovrapposte utilizzate

nei vari reparti, razionalizzando la spesa e semplificando la dotazione tecnologica.

Adobe ha costruito anche una piattaforma “autorigenerante” che individua problemi tecnici reali o emergenti (il tipo di guasti e difetti per cui un amministratore di sistema potrebbe ricevere una telefonata in piena notte) e li risolve automaticamente.

Secondo Stoddard, “per la nostra azienda l’AI è centrale per offrire esperienze ai nostri clienti. E, come organizzazione IT, seguiamo gli stessi principi per integrare in quello che facciamo funzionalità che semplifichino la vita dei nostri costituenti”.



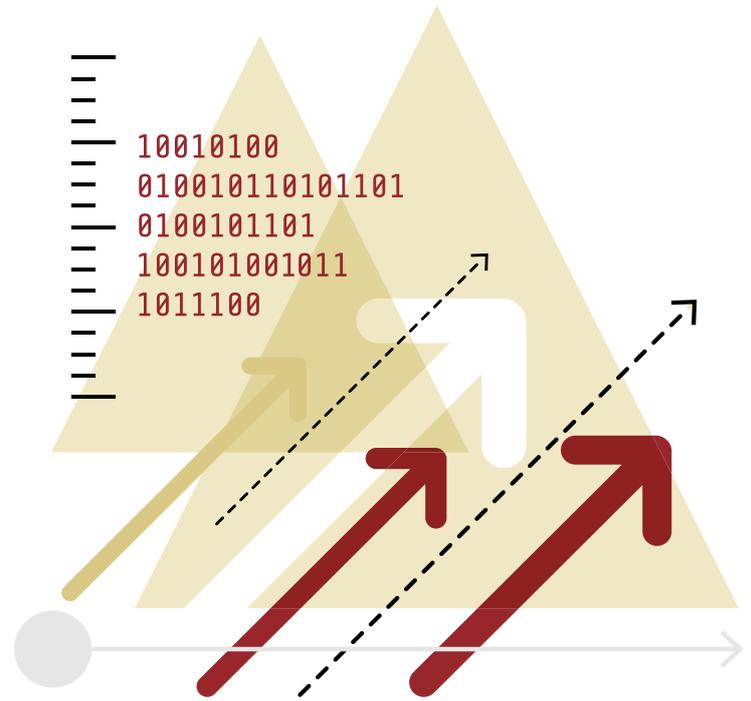
”Stiamo capendo la necessità di avere modelli di governance fortemente integrati, strutture di governance integrate per tutti i dati e tutti i modelli.”

Richard Spencer Schaefer,
Chief Health Informatics Officer,
Kansas City VA Medical Center

governance centralizzata: “Invece di fare tutto io direttamente, sto provando a rendere partecipe ogni utente e far capire a tutti che possono lavorare in autonomia. Responsabilizzando il team in questo modo, libero parte del mio tempo per concentrarmi sulla sicurezza e garantire la governance dei dati, in modo da evitare qualsiasi tipo di problema relativo ai dati”.

L'importanza della governance unificata per gestire i rischi dell'AI generativa è stato un argomento ricorrente nelle nostre interviste. Schaefer dice che al VA “stiamo capendo la necessità di avere modelli di governance fortemente integrati, strutture di governance integrate per tutti i dati e tutti i modelli. Stiamo prestando sempre maggiore attenzione a strumenti e processi estremamente centralizzati, che ci consentano di realizzare una struttura di dati aziendale”. E, se la governance unificata è sempre stata un'esigenza, l'AI generativa ha alzato la posta in palio. “Il rischio di avere dati non standardizzati e non ben definiti che girano in un modello, con tutte le conseguenze in termini di bias e deriva del modello stesso, ha reso questo aspetto molto più importante”, aggiunge.

Stoddard sottolinea l'importanza di ascoltare molte voci nel processo di valutazione dell'AI in Adobe: “È importante avere diverse valutazioni lungo tutto il processo e assicurarsi la necessaria diversità non solo in termini di etnia, genere e orientamento sessuale, ma anche di pensiero ed esperienza professionale che confluiscono nel processo e nella valutazione dell'impatto dell'AI”. Ed altrettanto importante conta anche la visibilità a tutti i livelli dell'organizzazione. Aggiunge Schaefer: “In cima alla nostra lista c'è l'adozione di strumenti di governance che offrano una panoramica visuale dei modelli in corso di



sviluppo, in modo che possano essere consultati dalla direzione o revisionati dalle figure interessate in qualsiasi momento”.

L'AI costituzionale, un approccio attualmente promosso dalla startup Anthropic, fornisce LLM con valori e principi specifici per aderire (invece di affidarsi) al feedback umano per guidare la produzione di contenuti²⁶. L'approccio costituzionale guida un modello affinché metta in atto le norme delineate nella costituzione, evitando ad esempio qualsiasi esito che possa essere tossico o discriminatorio.

Nonostante le tecnologie AI comportino rischi concreti per le aziende, d'altro canto esse offrono anche un grande potenziale per ridurre alcuni di questi rischi. Come osserva Zaharia, “analizzare i risultati dei modelli o analizzare il feedback in base ai commenti delle persone diventa più facile con i modelli linguistici. È quindi un po' più facile capire se il sistema sta commettendo qualche errore, e lo stiamo già osservando”. Il nostro sondaggio nel 2022 ha rilevato che la gestione della sicurezza e del rischio (31%) era il beneficio tangibile principale indicato dai responsabili AI, mentre il rilevamento delle frodi (27%), la sicurezza informatica (27%) e la gestione del rischio (26%) erano i primi tre impatti positivi previsti entro il 2025.

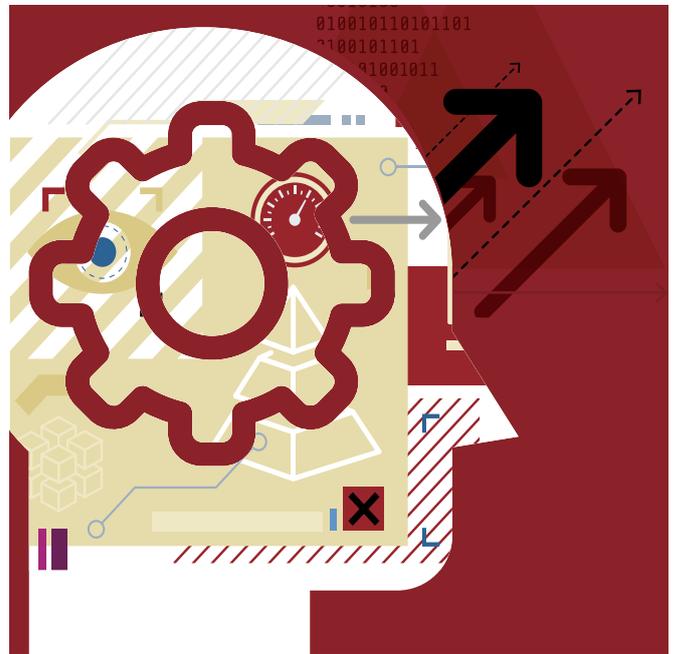
Una nuova tecnologia potente come l'AI generativa porta con sé numerosi rischi e responsabilità. Le nostre interviste indicano che una community AI motivata di professionisti, startup e aziende gestirà sempre più attentamente i rischi di governance dell'AI, così come fanno con la sostenibilità ambientale, attraverso una combinazione di preoccupazione per l'interesse pubblico, buona governance e tutela del marchio.

07 Conclusioni

AI ha attraversato molte primavere e inverni negli ultimi 50 anni. Fino a poco tempo fa, non si era fatta molta strada nelle aziende, al di là di progetti pilota o applicazioni in funzioni avanzate come IT e finanza. Questa situazione è destinata a cambiare nell'era dell'AI generativa, che sta favorendo l'adozione dell'intelligenza artificiale in tutta l'azienda.

Alcuni professionisti liquidano l'AI generativa come “moda del momento”. In realtà, i suoi critici sono impressionati dalla potenza dell'AI più di chiunque altro. Dirigenti ed esperti credono tuttavia che le organizzazioni possano cogliere i frutti e tenere a bada i rischi. Organizzazioni commerciali e governi devono entrambi muoversi lungo una linea sottile, accogliendo l'AI per accelerare l'innovazione e aumentare la produttività e, al tempo stesso, adottando le misure necessarie per contenere i rischi e prevedere gli incidenti che inevitabilmente accadranno.

Tenendo ben in mente queste precauzioni, i CIO più lungimiranti si stanno addentrando in modo deciso nell'era dell'AI. “Chi ha vissuto la rivoluzione del computer e di Internet ricorda quando arrivarono i primi computer”, dice Blyton. “Chi è stato fra i primi a imparare come lavorare con un computer, ha fatto una brillante carriera. Questa è una svolta paragonabile: se abbracci la tecnologia, ne trarrai sicuramente beneficio”.



“Chi è stato fra i primi a imparare come lavorare con un computer, ha fatto una brillante carriera. Questa è una svolta paragonabile: se abbracci la tecnologia, ne trarrai sicuramente beneficio”.

Andrew Blyton, Vice President
e Chief Information Officer,
DuPont Water & Protection

Informazioni su MIT Technology Review Insights

MIT Technology Review Insights è la divisione editoriale di *MIT Technology Review*, la più longeva rivista di tecnologie al mondo, sostenuta dalla più importante istituzione tecnologica mondiale, e produce eventi dal vivo e studi sulle principali tecnologie e sfide per il mondo delle imprese. Insights conduce ricerche e analisi qualitative e quantitative negli Stati Uniti e in altri Paesi, pubblicando un'ampia gamma di contenuti, fra cui articoli, rapporti, infografiche, video e podcast. Attraverso il MIT Technology Review [Global Insights Panel](#), Insights ha un accesso esclusivo ad alti dirigenti, innovatori e opinion leader di tutto il mondo, per sondaggi e approfondimenti.

Informazioni su Databricks

Databricks è un'azienda di AI e dati. Più di 10.000 organizzazioni in tutto il mondo (fra cui Comcast, Condé Nast e oltre il 50% delle aziende Fortune 500) fanno affidamento sulla piattaforma Databricks Lakehouse per unificare analisi dei dati e AI. Databricks ha la sede principale a San Francisco e uffici in tutto il mondo. Fondata dai creatori di Apache Spark™, Delta Lake e MLflow, Databricks persegue la missione di aiutare i team di gestione dei dati a risolvere i problemi più difficili del mondo. Nell'aprile 2023, Databricks ha rilasciato Dolly, il primo LLM open-source al mondo in grado di seguire istruzioni, ottimizzato su un dataset di istruzioni generate da esseri umani e disponibile con licenza per ricerca e uso commerciale. Per maggiori informazioni, segui Databricks su [Twitter](#), [LinkedIn](#) e [Facebook](#).



Note conclusive

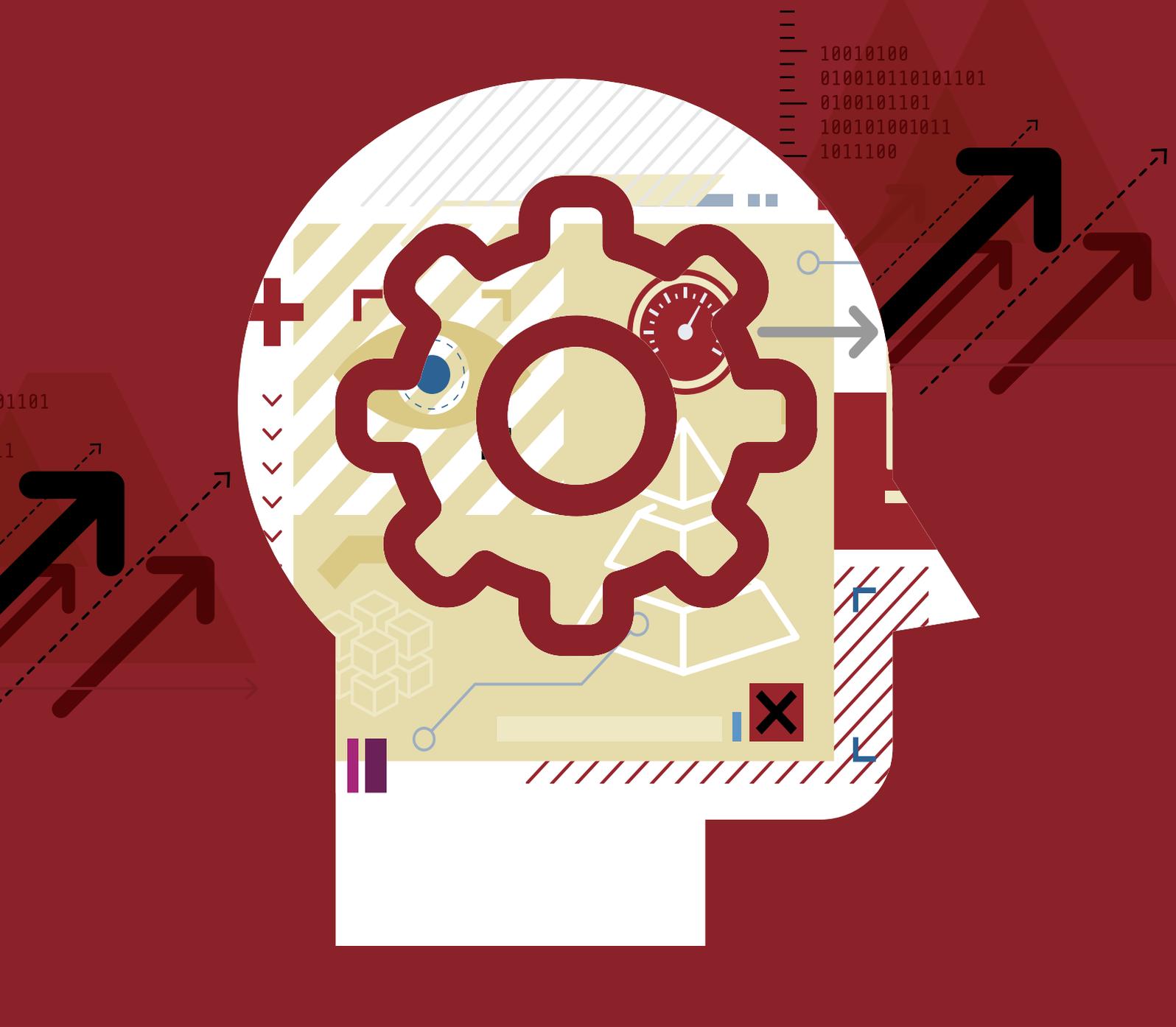
1. "The economic potential of generative AI", McKinsey & Company, 14 giugno 2023, <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#/>
2. "Generative AI could raise global GDP by 7%", Goldman Sachs, 5 aprile 2023, <https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html>
3. Jim Euchner, "Generative AI", Research-Technology Management, 20 aprile 2023, <https://doi.org/10.1080/08956308.2023.2188861>
4. "CIO vision 2025: Bridging the gap between BI and AI", MIT Technology Review Insights, settembre 2022, <https://www.technologyreview.com/2022/09/20/1059630/cio-vision-2025-bridging-the-gap-between-bi-and-ai/>
5. "Retail in the Age of Generative AI", Databricks, 13 aprile 2023, <https://www.databricks.com/blog/2023/04/13/retail-age-generative-ai.html>
6. "The Great Unlock: Large Language Models in Manufacturing", Databricks, 30 maggio 2023, <https://www.databricks.com/blog/great-unlock-large-language-models-manufacturing>
7. "Generative AI Is Everything Everywhere, All at Once", Databricks, 7 giugno 2023, <https://www.databricks.com/blog/generative-ai-everything-everywhere-all-at-once>
8. "Large Language Models in Media & Entertainment", Databricks, 6 giugno 2023, <https://www.databricks.com/blog/large-language-models-media-entertainment>
9. Ibid.
10. Ibid.
11. Ibid.
12. Ibid.
13. Jonathan Vanian and Kif Leswing, "ChatGPT and generative AI are booming, but the costs can be extraordinary", CNBC, 13 marzo 2023, <https://www.cnbc.com/2023/03/13/chatgpt-and-generative-ai-are-booming-but-at-a-very-expensive-price.html>
14. Josh Saul and Dina Bass, "Artificial Intelligence Is Booming – So Is Its Carbon Footprint", Bloomberg, 9 marzo 2023, <https://www.bloomberg.com/news/articles/2023-03-09/how-much-energy-do-ai-and-chatgpt-use-no-one-knows-for-sure#ix4y7vzkq>
15. Chris Stokel-Walker, "The Generative AI Race Has a Dirty Secret", Wired, 10 febbraio 2023, <https://www.wired.co.uk/article/the-generative-ai-search-race-has-a-dirty-secret>
16. Vanian and Leswing, "ChatGPT and generative AI are booming", <https://www.cnbc.com/2023/03/13/chatgpt-and-generative-ai-are-booming-but-at-a-very-expensive-price.html>
17. Arthur Spirling, "Why open-source generative AI models are an ethical way forward for science", Nature, 18 aprile 2023, <https://www.nature.com/articles/d41586-023-01295-4>
18. Sharon Goldman, "With a wave of new LLMs, open-source AI is having a moment – and a red-hot debate", VentureBeat, 10 aprile 2023, <https://venturebeat.com/ai/with-a-wave-of-new-llms-open-source-ai-is-having-a-moment-and-a-red-hot-debate/>
19. Thomas H. Davenport and Nitin Mittal, "How Generative AI Is Changing Creative Work", Harvard Business Review, 14 novembre 2022, <https://hbr.org/2022/11/how-generative-ai-is-changing-creative-work>
20. "Hello Dolly: Democratizing the magic of ChatGPT with open models", Databricks, 24 marzo 2023, <https://www.databricks.com/blog/2023/03/24/hello-dolly-democratizing-magic-chatgpt-open-models.html>
21. "Introducing the World's First Truly Open Instruction-Tuned LLM", Databricks, 12 aprile 2023, <https://www.databricks.com/blog/2023/04/12/dolly-first-open-commercially-viable-instruction-tuned-llm>
22. "Just how good can China get at generative AI?" The Economist, 9 maggio 2023, <https://www.economist.com/business/2023/05/09/just-how-good-can-china-get-at-generative-ai>
23. "Gen AI LLM – A new era of generative AI for everyone", Accenture, 17 aprile 2023, <https://www.accenture.com/content/dam/accenture/final/accenture-com/document/Accenture-A-New-Era-of-Generative-AI-for-Everyone.pdf>
24. "The economic potential of generative AI", McKinsey & Company, 14 giugno 2023, <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#/>
25. "Generative AI could raise global GDP by 7%", Goldman Sachs, 5 aprile 2023, <https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html>
26. "Claude's Constitution", Anthropic, 9 maggio 2023, <https://www.anthropic.com/index/claudes-constitution>

Illustrazioni

Tutte le illustrazioni sono ricavate da icone di Adobe Stock and The Noun Project.

L'accuratezza delle informazioni è stata verificata con la massima cura, tuttavia MIT Technology Review Insights declina qualsiasi responsabilità per la fiducia riposta da chiunque nel presente rapporto o in qualsiasi informazione, opinione o conclusione contenuta nel presente rapporto.

© Copyright MIT Technology Review Insights, 2023. Tutti i diritti riservati.



MIT Technology Review Insights

www.technologyreview.com
insights@technologyreview.com