

IL LIBRO DELL'INTELLIGENZA ARTIFICIALE

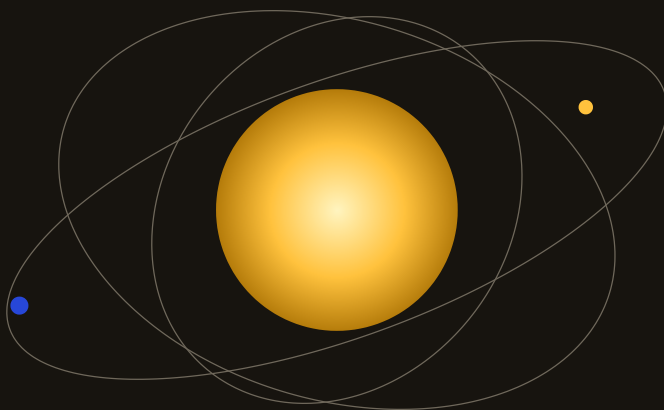
Lumen.

L'AI, spiegata
con la luce accesa.

Capire come funziona.

Scegliere come usarla.

Conservare il proprio giudizio.



Mattia Panatta

14 CAPITOLI · 5 PARTI
PRIMA EDIZIONE · OTTOBRE 2026

LUMEN · CAPIRE, USARE, DECIDERE

Lumen.

Il libro dell'intelligenza artificiale

Mattia Panatta

Quattordici capitoli, un percorso.

Modelli, strumenti, lavoro, responsabilità e futuro.

PRIMA EDIZIONE · OTTOBRE 2026
LUMENITALIA.COM

Il percorso del libro

Prima di cominciare · obiettivo e percorsi di lettura	4
---	---

PARTE I – CAPIRE

1. Settant'anni in una notte	9
2. Dentro la scatola	30
3. Pensare prima di rispondere	49

PARTE II – SCEGLIERE

4. La mappa dei modelli	70
5. La cassetta degli attrezzi	92
6. Agenti	111

PARTE III – USARE

7. Parlare con le macchine	130
8. L'AI al lavoro	147
9. Costruire con l'AI	164

PARTE IV – PROTEGGERSI

10. Quando l'AI sbaglia	183
11. Privacy, sicurezza, deepfake e truffe	199
12. Le regole del gioco	217

PARTE V – GUARDARE AVANTI

13. Lavoro, economia, energia	237
14. Le domande aperte	256

Epilogo · Il giudizio resta un lavoro	276	Appendice D — Scegliere senza inseguire la classifica	305
Appendice A — Ottanta parole per orientarsi	278	Appendice E — Aggiornarsi con un criterio	307
Appendice B — Cinquanta prompt da adattare	290	Quaderno operativo · sette materiali di lavoro	309
Appendice C — Quattro checklist da usare	302	L'autore e questa edizione	348

Prima di cominciare

Accendere la luce

Apri un assistente, scrivi una richiesta e ottieni un testo. Il gesto è semplice. Decidere se quel testo serva, se sia corretto e se tu possa usarlo richiede qualcosa in più. Da questo spazio fra facilità d'accesso e qualità del risultato nasce Lumen.

Sul sito la promessa è «L'AI, spiegata con la luce accesa»: rendere comprensibile una tecnologia che viene spesso raccontata per annunci, timori e classifiche. Questo libro raccoglie quella promessa in un percorso continuo. Ti accompagna dal funzionamento dei modelli alle scelte di strumenti, dall'uso quotidiano alle responsabilità, fino alle domande sul futuro.

Puoi leggerlo senza saper programmare. Servono curiosità e disponibilità a controllare una risposta anche quando suona bene. Se lavori in uno studio o in un'impresa, troverai un modo per delimitare le prove e misurare il lavoro completo. Se studi, troverai una distinzione a cui tornare: ottenere un esercizio svolto e imparare a svolgerlo sono risultati diversi. Se parti per interesse personale, puoi costruire un lessico con cui valutare le notizie senza dipendere da ogni nuova etichetta.

L'obiettivo è lasciarti più autonomo. Alla fine dovresti saper descrivere un compito, scegliere un ambiente adatto, fornire materiali utilizzabili, controllare il risultato e riconoscere quando occorre fermarsi. Non è necessario adottare ogni novità per raggiungere questo obiettivo.

Come è costruito il percorso

Le cinque parti seguono l'ordine di una decisione ragionata. **Capire** spiega da dove arrivano i sistemi e come producono risposte. **Scegliere** distingue modelli, prodotti e agenti. **Usare** mette il metodo alla prova nelle conversazioni, nel lavoro e nella costruzione di strumenti.

Proteggersi affronta errori, dati, sicurezza e regole. **Guardare avanti** collega il singolo uso a lavoro, economia, energia e questioni aperte.

Ogni capitolo dichiara che cosa ti permetterà di fare, sviluppa un ragionamento, propone schemi ed esempi e termina con un'attività. Le fonti stanno accanto alle affermazioni attraverso richiami numerati. Nel volume unico ogni richiamo porta alla nota del relativo capitolo; nell'HTML un collegamento di ritorno riporta al testo. Il PDF offre anche segnalibri per la consultazione. Un numero di nota si ripete in capitoli diversi: il collegamento resta distinto.

Le appendici sono strumenti di consultazione: **80 termini**, **50 prompt adattabili**, quattro checklist, una guida alle tabelle e alle prove di scelta, risorse per aggiornarsi. Il quaderno operativo conserva gli allegati già richiamati nel testo: schede compilabili, casi documentati, scenari di automazione e domande di verifica. Puoi usarli senza cercare altri file nella cartella del progetto.

Scegli da dove entrare

Se vuoi...	Parti da...	Poi metti alla prova...
Capire ciò che leggi nelle notizie	Capitoli 1–3	Una spiegazione del modello con parole tue, usando il glossario solo per i dubbi
Usare meglio un assistente	Capitoli 4, 5 e 7	Tre richieste dello stesso compito e un criterio per confrontarle

Se vuoi...	Parti da...	Poi metti alla prova...
Introdurre l'AI nello studio o nella PMI	Capitoli 7–8, poi 10–12	Un pilota delimitato; estendi soltanto se le misure lo giustificano
Costruire un'automazione	Capitoli 6 e 9, poi 10–12	Un caso ordinario, uno incompleto e uno che deve fermarsi
Studiare con più autonomia	Capitoli 7–8 e 14	Un esempio guidato seguito da un problema nuovo senza assistenza
Valutare conseguenze e futuro	Capitoli 12–14	Una tesi, la prova che la sostiene e ciò che potrebbe smentirla

Questi ingressi non eliminano i collegamenti fra le parti. Prima di consegnare a un sistema documenti riservati, consulta il capitolo 11. Prima di affidargli azioni, leggi i controlli dei capitoli 6 e 9. Puoi rimandare la storia; non puoi rimandare una verifica necessaria al compito che stai già svolgendo.

Il patto con il lettore

Fatti, esempi e opinioni hanno pesi diversi. Le fonti ufficiali identificano ciò che un ente o un fornitore dichiara; non trasformano automaticamente una promessa commerciale in una prova indipendente. Gli studi hanno popolazioni, compiti e versioni. I casi inventati servono per esercitarsi e non sono esperienze personali dell'autore. Quando leggi «secondo me» o una proposta operativa di Lumen, stai incontrando una valutazione argomentata, non un risultato sperimentale.

La data fa parte dell'informazione. Questa edizione fotografa testi e verifiche al **16 ottobre 2026**. Prezzi, accessi, modelli e regole possono cambiare. Un annuncio non garantisce una funzione sul tuo account; una norma pubblicata può avere una decorrenza successiva. Le note

conservano queste differenze e segnalano le informazioni non verificate. Prima di acquistare o prendere una decisione professionale, torna alle fonti pertinenti.

Il libro e il sito svolgono due lavori complementari. Il libro rende leggibile il filo del ragionamento; lumenitalia.com offre guide, confronti e opinioni che possono essere aggiornati. Una pagina online può quindi differire dalla fotografia conservata qui. Non assumere che tutti i prezzi del libro si aggiornino quando riapri l'HTML o il PDF: entrambi sono file di questa edizione.

Puoi imparare senza esporre documenti reali. Comincia con materiali pubblici, inventati o autorizzati e ridotti al necessario. I prompt in appendice richiedono adattamento: nessuna formula sostituisce la scelta dell'ambiente, le verifiche sui dati e il controllo umano. Il risultato utile è quello che riesci a capire, correggere e usare con responsabilità.

PARTE I

Capire

Dal percorso storico al funzionamento dei modelli. Una base per leggere capacità e limiti senza affidarsi alle metafore.

CAPITOLI 1-3

1. Settant'anni in una notte

La storia dell'intelligenza artificiale

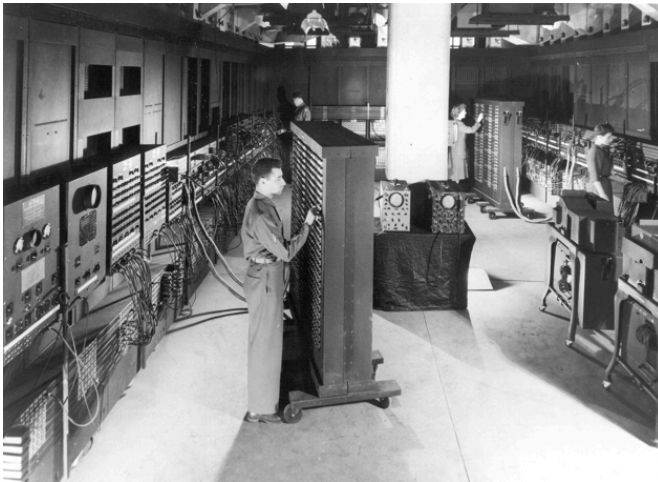
Lumen. Il libro dell'intelligenza artificiale · Mattia Panatta

PARTE I — CAPIRE · BASE · Circa 23 minuti di lettura

Testo e fonti verificati il 16 ottobre 2026. Le date degli annunci non implicano disponibilità immediata in ogni paese.

«Possiamo vedere solo un breve tratto davanti a noi, ma vediamo già molto che deve essere fatto.»

Alan M. Turing, *Computing Machinery and Intelligence*, 1950. Traduzione dall'inglese per questo capitolo.¹



ENIAC alla Moore School, fotografia del 1946. U.S. Army Photo · Wikimedia Commons. Dominio pubblico negli Stati Uniti.

In questo capitolo

- Da dove vengono le idee che oggi ritrovi nei chatbot.
 - Perché le promesse dell'AI hanno attraversato entusiasmi e delusioni.
 - Che cosa cambia con dati, reti neurali e Transformer.
 - Perché ChatGPT segna una svolta nell'accesso alla tecnologia.
 - Come si arriva ai modelli multimodali, al ragionamento e agli agenti del 2026.
-

Apri una chat, descrivi un problema e arriva una risposta. Puoi chiedere di riscriverla, allegare un documento, correggere un passaggio.

L'interazione è abbastanza semplice da far sparire tutto il lavoro che la rende possibile.

È anche per questo che l'intelligenza artificiale può sembrare comparsa all'improvviso. Per molte persone l'incontro ha una data: il 30 novembre 2022, quando OpenAI presenta ChatGPT al pubblico.² La ricerca che permette quell'incontro, però, attraversa decenni.

Nel titolo, i settant'anni vanno dall'incontro di Dartmouth del 1956 al 2026. La storia comincia ancora prima. La notte è quella della nostra percezione: un cambiamento preparato a lungo entra nella vita quotidiana quasi tutto insieme.

Conoscerne il percorso serve a valutare ciò che hai davanti. Ti aiuta a distinguere una capacità concreta dalla promessa che le viene costruita intorno, e a capire perché uno strumento può sorprenderti e sbagliare nella stessa conversazione.

1.1 Prima del nome, le idee

Nel 1943 Warren McCulloch, neurofisiologo, e Walter Pitts, studioso di logica, pubblicano un lavoro che mette in relazione attività nervosa e calcolo. Descrivono neuroni fortemente semplificati: unità che ricevono

segnali e si attivano quando sono soddisfatte certe condizioni. Collegandole, mostrano come si possano rappresentare operazioni logiche.³

È un modello matematico, molto lontano da una ricostruzione del cervello. Contiene però un'idea feconda: un comportamento complesso può emergere dall'interazione di unità semplici. Questi neuroni non imparano ancora come faranno le reti successive. La somiglianza con la biologia va presa per quello che è: un'ispirazione, non un'identità.

Nel 1950 Alan Turing affronta il problema da un'altra direzione. In *Computing Machinery and Intelligence* propone di sostituire una domanda difficile da definire — se una macchina possa pensare — con un confronto sul comportamento. Il suo gioco dell'imitazione parte da una situazione fra persone e viene riformulato introducendo una macchina. Da qui deriva ciò che chiamiamo test di Turing: valutare le risposte in un'interazione scritta, senza vedere chi le produce.¹

Quel confronto non misura direttamente coscienza o comprensione. Sposta l'attenzione verso qualcosa che possiamo osservare. È una distinzione utile anche davanti a un assistente moderno: una risposta convincente è un comportamento da esaminare; da sola non risolve la domanda su che cosa accada dentro il sistema.

Il nome arriva poco dopo. Il 31 agosto 1955 John McCarthy, Marvin Minsky, Nathaniel Rochester e Claude Shannon firmano la proposta del *Dartmouth Summer Research Project on Artificial Intelligence*. Il progetto prevede un lavoro estivo nel 1956, a Dartmouth, negli Stati Uniti. La formula *artificial intelligence*, associata a McCarthy, è dunque già nel documento del 1955.⁴

L'ambizione è ampia: capire se apprendimento e altri aspetti dell'intelligenza possano essere descritti abbastanza precisamente da essere simulati. La proposta immagina due mesi e un gruppo di dieci studiosi. Questo è il programma previsto, non un registro delle presenze effettive. Il seminario del 1956 diventerà il riferimento convenzionale per la nascita del campo.⁴

Intanto si sviluppa un'altra strada. Frank Rosenblatt lavora al **perceptron**, un sistema che modifica i propri collegamenti in base agli esempi. Nel 1958 una dimostrazione mostra un computer che impara a distinguere schede segnate a sinistra da schede segnate a destra. Il compito è ristretto; il cambiamento di prospettiva è grande: una parte del comportamento viene appresa, anziché descritta interamente dal programmatore.⁵

Da queste origini emergono due domande che accompagneranno tutta la storia: come rappresentare la conoscenza con regole, e come far apprendere una macchina dall'esperienza. Le risposte si svilupperanno in parallelo, con molti incroci.

Quattro nomi, ambiti diversi. AI è il campo più ampio. Il **machine learning**, o apprendimento automatico, comprende metodi che apprendono regolarità dai dati; il **deep learning** ne è una famiglia basata su reti a più strati. **Generativa** descrive la produzione di contenuti, come testo, immagini o musica. Un classificatore che distingue messaggi indesiderati e un generatore che scrive una risposta possono entrambi usare apprendimento automatico, ma svolgono lavori diversi.⁶ Un chatbot non rappresenta quindi tutta l'AI: nel libro ci concentreremo soprattutto sui sistemi generativi e sugli strumenti che li rendono utilizzabili.

Luce accesa — Il nome viene dopo l'idea

L'AI è un campo di ricerca che comprende metodi diversi. Un sistema che applica regole, uno che riconosce immagini e uno che genera testo possono appartenervi pur funzionando in modi molto differenti.

1.2 Le regole, le promesse e gli inverni

Una strada influente nei primi decenni è l'**AI simbolica**. La conoscenza viene espressa attraverso simboli, relazioni e regole che il programma manipola. Puoi immaginare una procedura per controllare una pratica: se

manca un documento obbligatorio, la pratica è incompleta; se è incompleta, occorre richiedere un'integrazione. È un'analogia didattica, ma rende visibile l'idea.

Il vantaggio è poter descrivere esplicitamente alcuni passaggi. La difficoltà arriva quando il mondo presenta situazioni che nessuno ha previsto, oppure quando le combinazioni possibili diventano troppe. Regole corrette, prese singolarmente, non garantiscono che l'intero problema sia gestibile.

Anche il linguaggio mette presto in luce lo scarto fra apparenza e funzionamento. Nel 1966 Joseph Weizenbaum pubblica ELIZA. Nella sua applicazione più nota, il programma riprende elementi delle frasi dell'interlocutore e li trasforma in domande, secondo uno stile ispirato al colloquio con uno psicoterapeuta. Si basa su riconoscimento di schemi e regole di trasformazione del testo.⁷

L'interazione può dare l'impressione di essere ascoltati. La lezione per noi è attuale: il modo in cui un sistema conversa influenza il giudizio che ne diamo. La sensazione di comprensione merita di essere distinta dalla verifica delle sue capacità.

Le aspettative crescono più rapidamente dei risultati utilizzabili. Il rapporto ALPAC del 1966 ridimensiona le attese sulla traduzione automatica negli Stati Uniti. Nel Regno Unito, il rapporto Lighthill del 1973 critica parte della ricerca sull'AI e i suoi risultati rispetto alle promesse, evidenziando anche la difficoltà di passare da problemi limitati a situazioni complesse. La critica viene contestata da ricercatori come McCarthy.^{8, 9}

Questi documenti sono passaggi di un raffreddamento più ampio. Il primo **inverno dell'AI** viene comunemente collocato fra la metà degli anni Settanta e l'inizio degli Ottanta. Il secondo, fra la fine degli anni Ottanta e i primi Novanta, segue il boom commerciale dei sistemi esperti. Sono periodizzazioni orientative: finanziamenti, fiducia e mercati si muovono in tempi diversi nei vari paesi.¹⁰

Fra i due periodi, i **sistemi esperti** offrono una via concreta all'applicazione. In un ambito delimitato, raccolgono conoscenze specialistiche e regole per formulare indicazioni. MYCIN, sviluppato a Stanford negli anni Settanta, è un esempio di ricerca sul supporto alla scelta delle terapie per alcune infezioni batteriche. La sua esistenza non equivale a un impiego clinico generalizzato.¹¹

Per costruire questi strumenti bisogna raccogliere il sapere degli esperti e tradurlo in una forma utilizzabile dal programma. Aggiornarlo diventa oneroso. I casi imprevisi espongono la fragilità delle regole, mentre l'acquisizione della conoscenza richiede molto lavoro: limiti discussi già dai ricercatori dell'epoca.¹² Alla disillusione contribuisce anche la crisi del mercato dell'hardware specializzato.¹⁰

Un inverno, quindi, non significa che ogni ricerca si fermi o che ogni applicazione perda valore. Significa che diventa più difficile sostenere certe aspettative con risultati e risorse. Per chi oggi valuta un progetto aziendale, è un richiamo concreto: una dimostrazione riuscita va seguita dalla prova su casi ordinari, eccezioni e costi di manutenzione.

1.3 Imparare dagli esempi diventa praticabile

Mentre il mercato attraversa queste oscillazioni, il lavoro sulle reti neurali continua. Una **rete neurale artificiale** collega unità di calcolo attraverso valori modificabili, chiamati pesi. Addestrarla significa regolare questi valori per migliorare il risultato su un compito.

Nel 1986 David Rumelhart, Geoffrey Hinton e Ronald Williams pubblicano un articolo fondamentale sulla **backpropagation**, o retropropagazione dell'errore. Il procedimento permette di calcolare come modificare i pesi nei diversi strati della rete a partire dall'errore del risultato. Il lavoro contribuisce in modo decisivo alla diffusione del metodo; non va presentato come un'invenzione senza antecedenti.¹³

Nel 1989 Yann LeCun e colleghi descrivono una rete applicata al riconoscimento delle cifre manoscritte dei codici postali. La struttura incorpora vincoli adatti alle immagini, anticipando la diffusione delle reti

convoluzionali: reti che cercano caratteristiche locali, come tratti e forme, riutilizzando gli stessi filtri in posizioni diverse.¹⁴

Intanto altre tecniche producono risultati spettacolari. Nel maggio 1997 Deep Blue di IBM batte Garry Kasparov, campione del mondo in carica, in un incontro di sei partite a tempo regolamentare: 3½ a 2½. Il sistema combina ricerca di molte posizioni e conoscenza scacchistica specialistica.¹⁵ La vittoria documenta una capacità precisa; non autorizza a estenderla automaticamente ad altri compiti.

Per le reti neurali diventano determinanti grandi raccolte di dati e maggiore capacità di calcolo. Le **GPU**, processori sviluppati per la grafica, sono adatte a eseguire in parallelo molte delle operazioni richieste dalle reti. Il **deep learning**, o apprendimento profondo, indica l'uso di reti con più strati, che apprendono rappresentazioni successive dei dati.

La svolta più visibile arriva nella competizione ImageNet del 2012. Alex Krizhevsky, Ilya Sutskever e Hinton presentano il lavoro associato ad **AlexNet**. La soluzione vincente combina più reti e raggiunge un errore *top-5* del 15,3%, contro il 26,2% del secondo classificato. “Top-5” significa che la categoria corretta deve comparire fra le cinque proposte. L'addestramento descritto nel lavoro usa due GPU.¹⁶

Il numero — 10,9 punti percentuali

È il divario fra i due errori top-5 appena citati, nella gara del 2012. È un confronto su un compito e un insieme di dati definiti: non una misura generale dell'intelligenza di AlexNet.¹⁶

Nel marzo 2016 AlphaGo di DeepMind batte Lee Sedol a Go per quattro partite a una. Combina reti neurali, ricerca delle mosse e apprendimento per rinforzo, nel quale un sistema migliora usando un segnale legato all'esito delle azioni. Nella seconda partita, la mossa 37 sorprende gli osservatori. Nella quarta, la mossa 78 di Lee contribuisce all'unica vittoria umana dell'incontro.¹⁷

Quelle due mosse rendono la storia più interessante di una semplice sostituzione dell'uomo. Un sistema può trovare soluzioni inattese e avere ancora punti deboli. La domanda utile diventa quali capacità abbia sviluppato e in quali condizioni funzionino.

La ricerca si estende anche alla scienza. Nel novembre 2020 DeepMind annuncia il risultato di AlphaFold 2 nella previsione della struttura delle proteine a CASP14, una valutazione internazionale. Predire strutture è un progresso importante, che non esaurisce la biologia delle proteine.¹⁸ Nel 2024 il Nobel per la chimica premia Demis Hassabis e John Jumper per la previsione strutturale, insieme a David Baker, premiato per la progettazione computazionale delle proteine.¹⁹

1.4 Il linguaggio diventa un punto di incontro

Riconoscere una cifra o scegliere una mossa sono compiti delimitati. Il linguaggio mette in relazione attività molto diverse: descrivere, spiegare, tradurre, formulare istruzioni. Per questo i progressi nella sua elaborazione acquistano un significato particolare.

Il 12 giugno 2017 otto ricercatori rendono pubblico *Attention Is All You Need*. Il lavoro presenta il **Transformer**, un'architettura basata sull'attenzione e sperimentata inizialmente soprattutto nella traduzione. Il contributo è un modo nuovo di organizzare componenti già oggetto di ricerca, con vantaggi nella parallelizzazione dell'addestramento.²⁰

L'**attenzione** permette di attribuire un peso diverso alle relazioni fra elementi di una sequenza. Non è attenzione cosciente. In una frase, aiuta a costruire rappresentazioni che tengono conto del contesto. I dettagli arriveranno nel capitolo 2; qui conta capire che un'architettura può cambiare la praticabilità di una strada di ricerca.

Nel 2018 OpenAI presenta il lavoro poi conosciuto come **GPT-1**. Un modello viene prima addestrato su molto testo e poi adattato a compiti specifici. L'acronimo GPT riassume *Generative Pre-trained Transformer*: un modello generativo basato su Transformer, con una fase di preaddestramento.²¹

Nello stesso anno il gruppo di Jacob Devlin presenta **BERT**, anch'esso basato sui Transformer. Durante il preaddestramento usa il contesto a sinistra e a destra per ricostruire parti nascoste del testo, ed è progettato per essere adattato a vari compiti di comprensione linguistica.²² I due lavori mostrano che un'idea architetturale può sostenere approcci differenti. GPT privilegia la generazione della continuazione; BERT costruisce rappresentazioni utili all'analisi del testo.

Con **GPT-2**, annunciato nel febbraio 2019, OpenAI mostra progressi nella generazione e sceglie inizialmente di non distribuire il modello completo, richiamando il rischio di usi impropri. La pubblicazione avviene per fasi.²³ Il rapporto fra diffusione della ricerca e controllo dei rischi entra così nella storia dei prodotti prima di ChatGPT.

Nel maggio 2020 il lavoro su **GPT-3** descrive un modello da 175 miliardi di parametri, cioè valori numerici regolati durante l'addestramento. Per diversi compiti bastano istruzioni o pochi esempi nel testo di ingresso, senza un nuovo aggiornamento dei pesi. È l'**apprendimento nel contesto**: il comportamento si adatta agli esempi presenti nella richiesta, senza che il modello debba per questo impararli permanentemente.²⁴

Nello stesso periodo, gli studi sulle **leggi di scala** descrivono relazioni empiriche fra dimensione dei modelli, dati, calcolo e miglioramento della previsione linguistica. Queste relazioni aiutano a pianificare esperimenti e investimenti; non dimostrano che ogni capacità cresca allo stesso ritmo o che ogni limite scompaia aumentando la scala.²⁵

Il passaggio decisivo è la possibilità di riutilizzare una base addestrata per molti compiti, anziché ripartire ogni volta da un sistema separato. Resta però una distanza fra produrre continuazioni linguistiche e comportarsi da assistente utile. Colmarla richiede altro lavoro.

Schema — Tre modi di costruire una risposta

Approccio	Da dove viene il comportamento	Che cosa devi verificare
Regole esplicite	Conoscenza e procedure descritte da persone	Copertura dei casi, coerenza e aggiornamento delle regole
Apprendimento dai dati	Regolarità apprese durante l'addestramento	Qualità dei dati e risultati su casi nuovi
Modello con strumenti	Capacità apprese, istruzioni e accesso a risorse esterne	Anche fonti consultate, azioni eseguite e permessi

Schema didattico: gli approcci possono convivere nello stesso sistema.

1.5 ChatGPT cambia la porta d'ingresso

Un modello che sa completare testi non segue necessariamente bene le tue istruzioni. Se gli presenti una domanda, potrebbe continuare un elenco di domande o imitare un documento, invece di risponderti nel modo desiderato.

Nel 2022 il lavoro su **InstructGPT** documenta un percorso per orientare i modelli verso le richieste delle persone. Si usano esempi di risposte e valutazioni umane per guidare l'addestramento. Una delle tecniche è il **reinforcement learning from human feedback**, abbreviato RLHF: apprendimento per rinforzo basato sul feedback umano.²⁶

L'obiettivo è rendere il comportamento più utile e più aderente alle istruzioni. Il gradimento umano, però, non coincide automaticamente con la verità: una risposta ben scritta può restare sbagliata. Questa distinzione accompagnerà tutto il libro.

Il **30 novembre 2022** OpenAI presenta ChatGPT come anteprima di ricerca gratuita. Il formato permette di fare domande successive, riprendere un errore e modificare la richiesta. Il post di lancio dichiara anche i limiti: il sistema può produrre risposte plausibili ma scorrette.²

La novità che incontra il pubblico è quindi un insieme di capacità, addestramento e interfaccia. Puoi cominciare scrivendo in italiano, senza conoscere il linguaggio di programmazione con cui sono stati costruiti i sistemi. Nella stessa conversazione puoi passare da una spiegazione a una bozza, poi chiedere una correzione.

Per documentare la diffusione senza confondere le metriche, basta un riferimento preciso: il **1° aprile 2024** OpenAI dichiara che oltre 100 milioni di persone usano ChatGPT ogni settimana. È un dato comunicato dall'azienda, relativo all'uso settimanale; non misura abbonamenti paganti o risparmi di tempo.²⁷ I numeri sulla popolarità vanno letti con la stessa cura dei risultati tecnici.

La storia, intanto, si allarga. Meta presenta LLaMA il 24 febbraio 2023, inizialmente con accesso destinato alla ricerca: non va confuso con una distribuzione commerciale priva di condizioni. Il 14 marzo arrivano gli annunci di GPT-4 e di Claude; il 6 dicembre Google presenta Gemini.²⁸
29, 30, 31

Per il lettore il cambiamento è anche pratico: la conversazione diventa un modo per accedere a strumenti di produttori diversi. Marchio, modello e applicazione rimangono cose distinte. Un nuovo nome sullo schermo può indicare un modello diverso, una funzione aggiunta o un prodotto costruito sopra capacità già esistenti. Impareremo a separarli nel capitolo 4.

Mito vs realtà

Mito: l'AI nasce con ChatGPT.

Realtà: ChatGPT rende accessibile a un pubblico ampio una combinazione maturata attraverso decenni di ricerca. Il lancio del 2022 è una svolta nell'interazione e nella diffusione, dentro una storia molto più lunga.

1.6 Dalle parole ai sensi, dal testo alle azioni

La generazione di contenuti percorre più strade. Nel 2014 Ian Goodfellow e colleghi presentano le **GAN**, reti generative avversarie: un generatore e un discriminatore vengono addestrati con obiettivi contrapposti. Nel 2020 il lavoro di Jonathan Ho, Ajay Jain e Pieter Abbeel contribuisce allo sviluppo dei **modelli di diffusione**, che imparano a ricostruire dati rimuovendo progressivamente il rumore.^{32, 33} Sono due tappe importanti, non un elenco completo dei metodi generativi.

Anche la musica entra in questo percorso. Il 16 novembre 2023 Google DeepMind presenta **Lyria**, insieme a esperimenti sviluppati con YouTube.³⁴ Per chi usa questi strumenti, il terreno si sposta: descrivere un risultato può diventare il punto di partenza per produrre materiali prima realizzati con procedure differenti.

I **modelli multimodali** lavorano con più tipi di contenuto, come testo, immagini e audio. Il 13 maggio 2024 OpenAI presenta **GPT-4o**, progettato per trattare queste modalità nello stesso modello. La presentazione e la distribuzione delle singole funzioni avvengono con tempi diversi.³⁵ Anche per **Sora** occorre distinguere le fasi: il rapporto tecnico viene pubblicato il 15 febbraio 2024; il prodotto Sora Turbo inizia a essere distribuito il 9 dicembre, con limitazioni geografiche al lancio.³⁶

Sono eventi storici, non indicazioni su ciò che puoi acquistare oggi. Uno strumento può essere modificato o ritirato: la stessa pagina di lancio di Sora segnala la cessazione del prodotto il 26 aprile 2026.³⁶ Per scegliere servizi disponibili servono verifiche aggiornate, alle quali sono dedicati i capitoli successivi.

Un altro cambiamento riguarda il tempo impiegato per rispondere. Il **12 settembre 2024** OpenAI annuncia **o1-preview**, una famiglia progettata per dedicare più elaborazione ai problemi prima della risposta. “Ragionamento” indica qui un comportamento e un’impostazione tecnica, senza risolvere la questione della somiglianza con il pensiero umano.³⁷

Il **20 gennaio 2025** DeepSeek pubblica **R1**, mettendo a disposizione pesi del modello e materiale tecnico. La pubblicazione amplia le possibilità di studiare e riutilizzare questo approccio.³⁸ Le affermazioni sulle prestazioni e sui costi richiedono confronti omogenei: non basta un annuncio per stabilire quanto costi riprodurre l'intero percorso di ricerca.

Poi cambia il rapporto con gli strumenti. Il **22 ottobre 2024** Anthropic presenta in beta pubblica la capacità di usare il computer; il **23 gennaio 2025** OpenAI presenta Operator come anteprima di ricerca.^{39, 40} Un sistema può osservare una schermata e proporre azioni su un'interfaccia. La riuscita su un compito non garantisce quella su tutti gli altri.

Un **agente**, nel senso pratico usato in questo libro, è un sistema che persegue un obiettivo attraverso più passaggi, usando strumenti e risultati intermedi. Il **29 settembre 2026** OpenAI presenta **dots** come agenti capaci di proseguire il lavoro nel tempo, con un proprio ambiente di calcolo. Per gli utenti Pro, la distribuzione iniziale esclude lo Spazio economico europeo, la Svizzera e il Regno Unito.⁴¹ "Presentato" non significa dunque "disponibile a tutti in Italia".

Fra i riferimenti recenti conserviamo l'annuncio di **Claude Opus 5.5 del 22 settembre 2026**, quello di **GPT-6.1 Sol al DevDay del 29 settembre** e la scheda di **Gemini 4 Argon aggiornata il 13 ottobre**. La data della scheda non dimostra quella del primo rilascio.^{42, 43, 44} Servono a collocare la fase storica: per il confronto fra offerte useremo la fotografia del capitolo 4.

La direzione è chiara: generare una risposta diventa una parte di sistemi che leggono materiali, elaborano problemi e possono intervenire su strumenti esterni. Valutarli richiede di osservare l'intero lavoro, compresi errori e controlli. Il capitolo 6 sarà dedicato a questa delega.

1.7 Una storia internazionale, una domanda concreta

La sequenza dei prodotti più visibili può far sembrare questa storia concentrata in pochi laboratori. Le persone e le istituzioni che la costruiscono attraversano invece paesi e discipline: logica, neuroscienze, informatica, statistica, ingegneria. Turing, LeCun e il lavoro di DeepMind sono già esempi di quanto il percorso europeo vi sia intrecciato.

Anche l'Italia ha una storia precedente ai chatbot. Nel **1973 Marco Somalvico fonda AIRLab al Politecnico di Milano**, laboratorio dedicato all'intelligenza artificiale e alla robotica.⁴⁵ La coincidenza con l'anno del rapporto Lighthill aiuta a leggere gli “inverni” con misura: il raffreddamento di alcuni programmi e finanziamenti convive con iniziative e continuità della ricerca altrove.

Questo percorso non consegna un metodo destinato a vincere sempre. Ci sono problemi per i quali regole esplicite e controlli deterministici restano adatti, e altri per i quali l'apprendimento dai dati apre possibilità nuove. Nei sistemi applicati possono convivere componenti differenti. La scelta utile parte dal compito e dal modo di verificarne l'esito.

Dal lavoro di studio — Un esempio didattico

Immagina di dover controllare un fascicolo. Una procedura può verificare se contiene i documenti previsti; un modello può aiutarti a leggere quelli presenti e preparare una sintesi; un agente può cercare i materiali nelle cartelle autorizzate e organizzare i passaggi. Sono responsabilità diverse, con controlli diversi. Questo è un esempio costruito per il libro, non il resoconto di un incarico realmente svolto dall'autore.

Per capire quanto fidarti di una nuova funzione, puoi ripercorrere in piccolo le domande di questa storia. Quale problema risolve? Su quali casi è stata provata? Che cosa accade quando il caso cambia? Chi verifica il risultato prima che abbia conseguenze?

Una vittoria a scacchi, un buon punteggio nel riconoscimento delle immagini e una conversazione fluida dimostrano capacità differenti. Anche un modello riutilizzabile su molti compiti va valutato nei singoli impieghi. La storia invita a riconoscere i progressi senza attribuire loro una portata che le prove non sostengono.

Il criterio che propongo in questo libro è giudicare l’AI dal lavoro che riesce a portare a termine con un livello di affidabilità adeguato allo scopo. Nel tempo complessivo vanno compresi preparazione, verifica e correzioni. È una scelta di metodo: ci accompagnerà quando parleremo di studio, professioni e imprese.

Con questo criterio puoi tornare alla chat iniziale e vederla meglio. Dietro la semplicità dell’interfaccia ci sono scelte su dati, addestramento e strumenti. Nel prossimo capitolo entreremo nel funzionamento di un modello linguistico: che cosa riceve quando scrivi, come costruisce una risposta e da dove vengono alcuni dei suoi errori.

La linea del tempo

Le date dei lavori scientifici indicano la prima pubblicazione citata; quelle dei prodotti, l’annuncio o la fase di distribuzione specificata. Gli intervalli degli inverni sono orientativi.

Dalle origini alle reti profonde

Quando	La tappa	Che cosa cambia
1943	McCulloch e Pitts ³	Un modello formale collega neuroni semplificati e logica.
1950	Il saggio di Turing ¹	Il comportamento diventa terreno di confronto.
1955 → 1956	Proposta e seminario di Dartmouth ⁴	Un nome e un progetto di ricerca comune.

Quando	La tappa	Che cosa cambia
1958	Il perceptron di Rosenblatt ⁵	Un sistema apprende una distinzione dagli esempi.
1966	ELIZA e rapporto ALPAC ^{7, 8}	Il dialogo simulato e i limiti della traduzione.
1973	Rapporto Lighthill; fondazione di AIRLab ^{9, 45}	Critiche alla ricerca e nuove iniziative coesistono.
Metà anni Settanta → primi Ottanta	Primo inverno ¹⁰	Si ridimensionano aspettative e sostegno in parte del campo.
Anni Settanta e Ottanta	Sistemi esperti ^{11, 12}	La conoscenza specialistica viene tradotta in regole.
1986 → 1989	Backpropagation e riconoscimento di cifre ^{13, 14}	L'apprendimento in reti a più strati trova applicazioni.
Fine anni Ottanta → primi Novanta	Secondo inverno ¹⁰	Emergono fragilità e costi del boom commerciale.
1997	Deep Blue ¹⁵	Un sistema batte il campione in carica nel match di scacchi.
2012	AlexNet a ImageNet ¹⁶	Dati, reti profonde e GPU producono un forte miglioramento.
2016	AlphaGo contro Lee Sedol ¹⁷	Apprendimento e ricerca delle mosse si combinano con successo.

Dal linguaggio agli agenti

Quando	La tappa	Che cosa cambia
2017	Transformer ²⁰	Una nuova architettura facilita l'addestramento parallelo.
2018 → 2020	GPT, BERT, GPT-2 e GPT-3 ^{21, 22, 23, 24}	Il preaddestramento linguistico alimenta molti compiti.
2020	Leggi di scala; AlphaFold ^{25, 18}	Si studia la crescita dei modelli; l'AI avanza nella scienza.
2022	InstructGPT; ChatGPT il 30 novembre ^{26, 2}	I modelli vengono orientati verso istruzioni e dialogo.
2023	LLaMA, GPT-4, Claude e Gemini ^{28, 29, 30, 31}	Si amplia il panorama dei modelli e delle applicazioni.
2023 → 2024	Lyria, GPT-4o e Sora ^{34, 35, 36}	Musica, audio, immagini e video acquistano nuove interfacce.
Settembre 2024 → gennaio 2025	o1-preview e DeepSeek-R1 ^{37, 38}	Si sviluppano nuove famiglie dedicate al ragionamento.
Ottobre 2024 → gennaio 2025	Computer use e Operator ^{39, 40}	L'interazione con le interfacce diventa uno strumento degli agenti.
Settembre → ottobre 2026	Opus 5.5, GPT-6.1 Sol, dots e documentazione di Gemini 4 Argon ^{41, 42, 43, 44}	Prosegue l'integrazione fra modelli, strumenti e lavoro continuativo.

In una riga

L'AI che incontri oggi nasce da decenni di idee, prove e correzioni: conoscerli ti aiuta a giudicare ciò che sa fare, oltre l'effetto della prima dimostrazione.

Da fare questa settimana

- ☐ **Ricostruisci tre passaggi.** Scegli Dartmouth, ImageNet o ChatGPT: apri la fonte indicata e annota che cosa viene effettivamente proposto o dimostrato.
- ☐ **Confronta due richieste.** Chiedi a un assistente di spiegare un concetto che conosci e di inventare un esempio. Valuta separatamente correttezza della spiegazione e utilità dell'esempio.
- ☐ **Esamina un annuncio.** Individua data, capacità dichiarata, prove e limiti di accesso. Se un'informazione manca, registrala come mancante.

Prova tu — Leggere un annuncio con la luce accesa

«Analizza il testo pubblico che ti fornisco. Separa le capacità dichiarate dai risultati misurati e dai programmi futuri. Riporta data, condizioni delle prove e limiti di disponibilità, quando presenti. Per ogni punto indica il passaggio che lo sostiene. Non aggiungere informazioni che il testo non contiene.»

Controlla poi l'analisi sul documento originale: anche un assistente incaricato di verificare può sbagliare.

Approfondisci su Lumen

Breve storia dell'intelligenza artificiale · Come funziona davvero un LLM.

Le guide accompagnano il libro e possono essere aggiornate nel tempo. Le fonti che sostengono questo capitolo sono riportate qui sotto.

Note e fonti

Consultazione e verifica: 16 ottobre 2026. Le fonti aziendali documentano annunci, caratteristiche dichiarate e condizioni di lancio; non costituiscono una valutazione indipendente delle prestazioni. Le ricostruzioni storiche sono distinte dai documenti originali.

Titoli sottolineati: link diretti alle fonti. Su Lumen: [Breve storia dell'intelligenza artificiale](#) · [Sito ufficiale](#) ⁷.

1. A. M. Turing, *Computing Machinery and Intelligence*, *Mind*, 59 (236), pp. 433–460, ottobre 1950. [Testo integrale ospitato dalla Michigan State University](#). Gioco dell'imitazione nelle sezioni iniziali; epigrafe dall'ultima frase, tradotta per questo capitolo.
2. OpenAI, *Introducing ChatGPT*, 30 novembre 2022. [Annuncio originale, metodo e limiti dichiarati](#). Il servizio viene presentato come anteprima di ricerca gratuita; non si attribuisce al lancio una garanzia di correttezza.
3. W. S. McCulloch, W. Pitts, *A Logical Calculus of the Ideas Immanent in Nervous Activity*, *Bulletin of Mathematical Biophysics*, 5, pp. 115–133, 1943. [Riproduzione del lavoro presso Carnegie Mellon University](#). Il file riproduce il testo storico con una successiva nota editoriale.
4. J. McCarthy, M. L. Minsky, N. Rochester, C. E. Shannon, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 31 agosto 1955. [Documento originale nell'archivio di McCarthy a Stanford](#). La proposta riguarda l'estate 1956; durata e numero di studiosi sono quelli programmati.
5. M. Lefkowitz, *Professor's perceptron paved the way for AI — 60 years too soon*, Cornell University, 25 settembre 2019. [Ricostruzione istituzionale e materiali d'archivio](#). Il 1958 è l'anno della dimostrazione qui descritta, non l'inizio di tutto il lavoro sul perceptron.
6. Google for Developers, [What is Machine Learning?](#), consultato il 16 ottobre 2026. Classificazione, previsione e generazione come compiti differenti; mappa introduttiva del capitolo, non equivalenza fra tutto il campo AI e i chatbot.
7. J. Weizenbaum, *ELIZA — A Computer Program for the Study of Natural Language Communication Between Man and Machine*, *Communications of the ACM*, 9 (1), pp. 36–45, gennaio 1966, [DOI del lavoro originale](#). Per la descrizione del funzionamento, anche [materiale didattico di Stanford su ELIZA](#). Il testo integrale ACM non è stato recuperato in questa verifica.
8. Automatic Language Processing Advisory Committee, *Language and Machines — Computers in Translation and Linguistics*, National Academy of Sciences / National Research Council, 1966. [Rapporto ALPAC presso l'editore istituzionale](#).
9. J. McCarthy, *Review of "Artificial Intelligence: A General Survey" by Professor Sir James Lighthill*, risposta al rapporto dello Science Research Council del 1973, conservata nell'archivio dell'autore; la pagina non indica una data autonoma di pubblicazione della recensione. [Testo della risposta e riferimento al rapporto](#). Fonte contemporanea di parte, utile anche per documentare il dissenso; per l'inquadramento storico, fonte 10.
10. T. E. Senator, *Implications for AI Research: Applying Lessons from the Expert Systems Boom and Bust to the Current Large-Language Model Boom*, *Proceedings of the AAAI Conference on Artificial Intelligence*, 40 (46), pp. 39780–39784, 14 marzo 2026. [Articolo](#) · [Testo integrale](#). Ricostruzione storica di un ricercatore, usata per contesto e periodizzazione; le analogie con il presente restano interpretazioni.
11. J. McCarthy, *Some Expert Systems Need Common Sense*, 1984. [Testo dell'autore a Stanford](#). Discute MYCIN e i limiti della conoscenza rappresentata nei sistemi esperti.
12. D. B. Lenat, M. Prakash, M. Shepherd, *CYC: Using Common Sense Knowledge to Overcome Brittleness and Knowledge Acquisition Bottlenecks*, *AI Magazine*, 6 (4), 1985. [Articolo e metadati AAAI](#). Si usa l'anno del volume: il fascicolo è indicato come "Winter 1985", mentre il sito espone anche una data di pubblicazione del 15 marzo 1985.
13. D. E. Rumelhart, G. E. Hinton, R. J. Williams, *Learning representations by back-propagating errors*, *Nature*, 323, pp. 533–536, 9 ottobre 1986. [Articolo e metadati dell'editore](#). Il riferimento bibliografico a Le Cun (1985), presente nel lavoro, è fra gli elementi che impediscono di descrivere il 1986 come origine senza precedenti.

- .4. Y. LeCun e altri, **Backpropagation Applied to Handwritten Zip Code Recognition**, *Neural Computation*, 1 (4), pp. 541–551, 1° dicembre 1989. [Abstract e metadati MIT Press](#). Verifica sul sommario pubblico; il PDF dell'editore non è stato recuperato.
- .5. IBM, **Deep Blue**, pagina storica istituzionale senza data editoriale esplicita nella versione consultata. [Ricostruzione dell'incontro del maggio 1997](#). Per l'11 maggio come data conclusiva, anche [presentazione IBM del 2007](#), [apertura](#).
- .6. A. Krizhevsky, I. Sutskever, G. E. Hinton, **ImageNet Classification with Deep Convolutional Neural Networks**, *Advances in Neural Information Processing Systems*, 25, 2012. [Paper originale](#), abstract e § 6, tabella 2. Il 15,3% è il risultato top-5 della combinazione di sette reti nella gara 2012. La pagina HTML dell'abstract contiene valori diversi relativi ad altri risultati: per questi numeri si è seguito il PDF.
- .7. Google DeepMind, **AlphaGo**, pagina di ricerca senza data editoriale esplicita. [Tecnica, incontro del marzo 2016, mosse 37 e 78](#). La pagina descrive il risultato di 4–1; le mosse appartengono rispettivamente alla seconda e alla quarta partita.
- .8. Google DeepMind, **AlphaFold: a solution to a 50-year-old grand challenge in biology**, 30 novembre 2020. [Annuncio dei risultati CASP14](#). Si distingue la previsione della struttura dalla spiegazione completa del comportamento delle proteine.
- .9. Royal Swedish Academy of Sciences / NobelPrize.org, **The Nobel Prize in Chemistry 2024**, 9 ottobre 2024. [Comunicato e motivazioni](#) · [Scheda di Demis Hassabis](#). Metà premio a Baker per la progettazione computazionale, metà congiuntamente a Hassabis e Jumper per la previsione strutturale.
- .10. A. Vaswani e altri, **Attention Is All You Need**, prima pubblicazione arXiv il 12 giugno 2017, poi *NeurIPS 2017*. [Paper e storia delle versioni](#). La data indicata è quella della prima versione, non della revisione del 2023.
- .11. OpenAI, **Improving language understanding with unsupervised learning**, 11 giugno 2018, con il lavoro di A. Radford e altri **Improving Language Understanding by Generative Pre-Training**. [Annuncio e collegamento al paper](#). “GPT-1” è la denominazione retrospettiva usata nel capitolo.
22. J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**, prima pubblicazione l'11 ottobre 2018. [Paper originale e versioni](#).
23. OpenAI, **Better language models and their implications**, 14 febbraio 2019. [Annuncio di GPT-2 e motivazioni della distribuzione iniziale limitata](#).
24. T. B. Brown e altri, **Language Models are Few-Shot Learners**, prima pubblicazione il 28 maggio 2020. [Paper su GPT-3](#). Dimensione di 175 miliardi di parametri e prove senza aggiornamento dei pesi per i compiti presentati nel contesto.
25. J. Kaplan e altri, **Scaling Laws for Neural Language Models**, prima pubblicazione il 23 gennaio 2020. [Paper e ambito delle relazioni empiriche](#).
26. L. Ouyang e altri, **Training language models to follow instructions with human feedback**, prima pubblicazione il 4 marzo 2022. [Paper su InstructGPT](#).
27. OpenAI, **Start using ChatGPT instantly**, 1° aprile 2024. [Dichiarazione di oltre 100 milioni di utenti settimanali](#). Dato aziendale di utilizzo, distinto da utenti mensili, account registrati e abbonati paganti.
28. Meta, **Introducing LLaMA: A foundational, 65-billion-parameter large language model**, 24 febbraio 2023. [Annuncio e condizioni iniziali di accesso](#).
29. OpenAI, **GPT-4**, 14 marzo 2023. [Annuncio di ricerca](#). Modello multimodale con testo e immagini in ingresso; non tutte le capacità sono distribuite simultaneamente.
30. Anthropic, **Introducing Claude**, 14 marzo 2023. [Annuncio originale](#).
31. Google, **Introducing Gemini: our largest and most capable AI model**, 6 dicembre 2023. [Annuncio originale](#).
32. I. J. Goodfellow e altri, **Generative Adversarial Networks**, prima pubblicazione il 10 giugno 2014. [Paper originale](#).
33. J. Ho, A. Jain, P. Abbeel, **Denosing Diffusion Probabilistic Models**, prima pubblicazione il 19 giugno 2020. [Paper originale](#). È una tappa dei modelli di diffusione, non la loro prima formulazione storica.
34. Google DeepMind, **Transforming the future of music creation**, 16 novembre 2023. [Annuncio di Lyria e degli esperimenti con YouTube](#).

35. OpenAI, Hello GPT-4o, 13 maggio 2024. [Architettura multimodale e distribuzione per fasi](#).
36. OpenAI, Video generation models as world simulators, 15 febbraio 2024, [rapporto tecnico su Sora](#); Sora is here, 9 dicembre 2024, [lancio di Sora Turbo](#). Quest'ultima pagina, aggiornata successivamente, segnala la cessazione del prodotto il 26 aprile 2026. Il capitolo distingue ricerca, lancio e successivo stato del servizio.
37. OpenAI, Introducing OpenAI o1-preview, 12 settembre 2024. [Annuncio originale](#). È una tappa della commercializzazione dei modelli di ragionamento, non l'origine di ogni ricerca sul ragionamento automatico.
38. DeepSeek, DeepSeek-R1 Release, 20 gennaio 2025. [Annuncio, licenza e collegamenti tecnici](#). Le dichiarazioni comparative dell'azienda non sono assunte come giudizio indipendente.
39. Anthropic, Introducing computer use, a new Claude 3.5 Sonnet, and Claude 3.5 Haiku, 22 ottobre 2024. [Annuncio della beta pubblica](#).
40. OpenAI, Introducing Operator, 23 gennaio 2025. [Annuncio dell'anteprima di ricerca](#). Si descrive la tappa storica, non lo stato attuale del prodotto.
41. OpenAI, Introducing dots, 29 settembre 2026. [Annuncio e funzionamento dichiarato](#); [ChatGPT release notes, voce del 29 settembre 2026 "Meet your dot"](#). Le note di rilascio precisano l'esclusione iniziale di SEE, Svizzera e Regno Unito per gli utenti Pro; distinguono l'accesso Enterprise in beta. Il capitolo non presume una disponibilità italiana successiva.
42. Anthropic, Introducing Claude Opus 5.5, 22 settembre 2026. [Annuncio ufficiale](#). Si verifica l'annuncio, senza adottare le valutazioni promozionali sulle prestazioni.
43. OpenAI, DevDay 2026 Recap, 29 settembre 2026. [Resoconto ufficiale del DevDay](#) e [annuncio di GPT-6.1 Sol](#). La data è verificata sul resoconto datato dell'evento e sulle note di rilascio, non dedotta da un elenco di modelli correnti.
44. Google DeepMind, Gemini 4 Argon — Model Card, aggiornata il 13 ottobre 2026. [Scheda tecnica](#) · [Registro delle schede e delle date](#). La data del registro è di aggiornamento documentale.
45. Politecnico di Milano, DEIB, Laboratorio AIRLab — Artificial intelligence and robotics, pagina istituzionale senza data editoriale esplicita. [Storia e attività del laboratorio](#). Documenta la fondazione da parte di Marco Somalvico nel 1973.