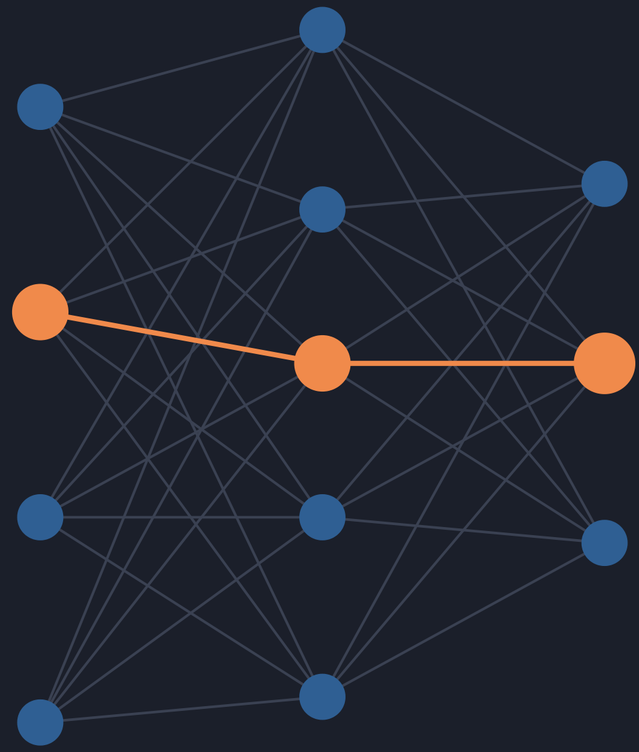
L’AI ha sempre una risposta. Ma è quella giusta?

Fiducia, scelte e pensiero critico nell'era dell'intelligenza artificiale



Gianluca Garofalo

Il percorso di oggi

INTRO

Cos'è l'AI generativa

Da dove viene, come funziona e cosa non è

1

Ha sempre una risposta

Come cambia il rapporto con la conoscenza

2

Perché ci fidiamo

Come cambia la fiducia nell'era dell'AI

3

Conosce il mondo?

Dati, bias e rappresentazioni della realtà

4

Chiedere o delegare

Tra farsi aiutare e smettere di decidere

5

Sbaglia e convince

Convincente non vuol dire corretta

6

Senza smettere di pensare

Tre domande davanti a una risposta

7

Vita quotidiana

Casa, scuola, lavoro e i nostri dati

8

Avere paura?

Rischi reali, opportunità e falsi allarmi

9

E noi?

Quale umanità vogliamo costruire

Dalle origini agli inverni dell’AI

1950

L'IDEA

Alan Turing si chiede se le macchine possano pensare

1956

IL NOME

A Dartmouth nasce il termine «intelligenza artificiale». Grandi promesse

1974–80

PRIMO INVERNO

Computer lenti e pochi dati: i risultati non arrivano e i fondi vengono tagliati

anni ’80

SISTEMI ESPERTI

Le aziende adottano l'AI a regole: migliaia di istruzioni «se… allora»

1987–93

SECONDO INVERNO

Regole fragili e costose da mantenere: il mercato crolla

La lezione degli inverni. Ogni volta che le promesse hanno superato i risultati, entusiasmo e investimenti sono crollati. Per funzionare davvero, l'AI aveva bisogno di più dati, più potenza di calcolo e metodi diversi.

Dalla rinascita al Transformer

anni ’90

LA RIPRESA

L'AI torna a crescere, senza clamore, con metodi statistici che imparano dai dati

1997

DEEP BLUE

IBM batte Kasparov a scacchi, ancora con regole e forza di calcolo

2012

DEEP LEARNING

Con tanti dati e le schede grafiche, le reti neurali riconoscono le immagini

2017

TRANSFORMER

Google presenta l'architettura alla base dei modelli linguistici attuali

Il Transformer, in breve. Legge tutta la frase in una volta e, per ogni parola, valuta quanto contano tutte le altre: è il meccanismo di «attenzione». In «Il cane non ha attraversato la strada perché era stanco», capisce che a essere stanco è il cane, non la strada.

Gli ultimi quattro anni

2022

ChatGPT porta l'AI generativa nelle mani di tutti

2023

Modelli multimodali: testo e immagini nello stesso modello

2024

Modelli che «ragionano»: elaborano passaggi intermedi prima di rispondere

2024

Entra in vigore l'AI Act, la prima legge europea sull'AI

2025–26

Agenti: l'AI pianifica, usa strumenti e porta a termine compiti

Che cos'è un agente. Un sistema costruito attorno a un modello: riceve un obiettivo invece di una domanda, lo scompone in passi, usa strumenti come il browser, i file o altri software, controlla il risultato e prosegue fino alla fine del compito.

Cos’è davvero l’AI

Esistono due grandi famiglie: l'AI che applica regole scritte da esperti e l'AI che impara dai dati.

AI DETERMINISTICA (A REGOLE)

Stesso input, stessa risposta. Sempre.

Ragiona applicando regole e conoscenze codificate da esperti umani. Prevedibile e spiegabile, ma sa solo ciò che le è stato insegnato in modo esplicito.

Sistemi esperti, Deep Blue, calcolo del percorso nel navigatore

MODELLI GENERATIVI

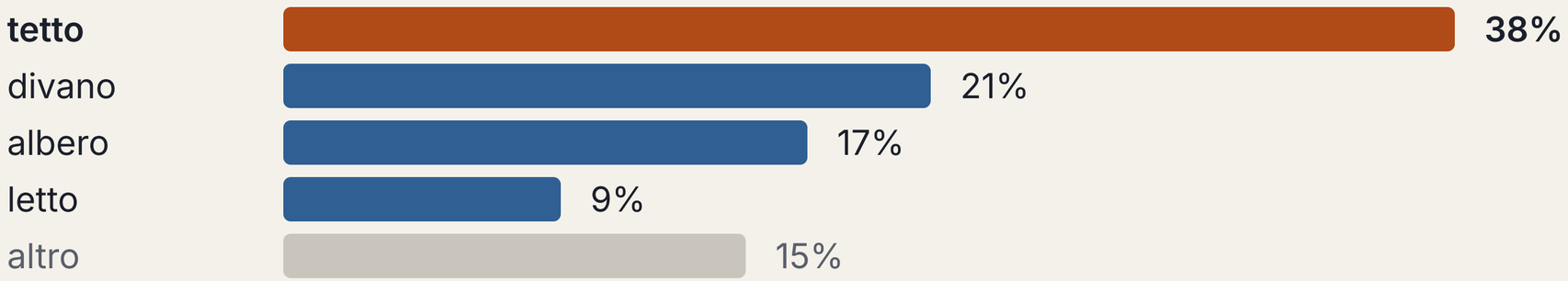
Stesso input, risposte diverse.

Ricavano schemi da enormi quantità di dati e producono contenuti nuovi scegliendo in base a probabilità. Flessibili, ma non garantiti.

ChatGPT, Claude, Gemini, generatori di immagini

Come funziona: una parola alla volta

«Il gatto è salito sul …»



1 Legge il testo fin qui

2 Calcola le probabilità

3 Sceglie una parola

4 La aggiunge e ricomincia

Token: l'unità su cui lavora davvero il modello. Può essere una parola intera, una parte di parola, una singola lettera o un segno di punteggiatura. Qui, per semplicità, diciamo «parola».

Percentuali illustrative, non misurate su un modello reale

E soprattutto, cosa non è

Non è una mente

Non pensa e non prova emozioni come una persona: elabora testo in base alle probabilità.

Non è un'enciclopedia

Non consulta un archivio di fatti verificati: costruisce ogni risposta parola dopo parola.

Non ha intenzioni proprie

Non vuole nulla da sé: persegue gli obiettivi che le assegniamo noi.

Non sa cosa è vero

Sa cosa è probabile. Spesso coincide con il vero, ma non sempre.

Dal cercare al chiedere

Con un motore di ricerca

- Tanti risultati da confrontare
- Le fonti sono in vista
- La scelta tocca a noi

Con un assistente AI

- Una sola risposta, già pronta
- Le fonti spesso non si vedono
- Risponde sempre, anche quando non sa

Guadagniamo tempo. Il rischio è perdere l'abitudine a confrontare e a scegliere.

Quattro motivi per cui ci fidiamo

Parla come noi

Un linguaggio fluente e ordinato ci sembra un segno di competenza.

Non esita mai

Usa lo stesso tono sicuro per le risposte giuste e per quelle sbagliate.

È sempre disponibile

Paziente, veloce, a qualsiasi ora: non si stanca e non giudica.

Tende a darci ragione

È addestrata a soddisfarci, e spesso asseconda quello che pensiamo già.

Prima ci fidavamo di fonti con un nome e un volto; ora di un sistema che non ne ha. **La fiducia va calibrata:** proporzionata a quanto possiamo verificare.

Come impara: dagli esempi

Come si insegna a una macchina a riconoscere un gatto, senza descriverlo?

1 · Mostrare

Migliaia di foto con l'etichetta giusta: gatto, non gatto.

2 · Sbagliare

Il modello tira a indovinare e misura quanto si è allontanato dalla risposta giusta.

3 · Correggere

Aggiusta di poco i suoi parametri, per sbagliare un po' meno la volta dopo.

Ripetuto milioni di volte, il modello scopre da solo quali forme contano. La regola «gatto» non l'ha scritta nessuno.

Come nasce un modello linguistico

1	**Raccolta dei dati**	Si raccolgono testi, immagini, codice e altri contenuti per l'addestramento.	**Misto** · persone scelgono le fonti, software le raccoglie
2	**Preparazione dei dati**	Si eliminano duplicati, contenuti indesiderati e dati di scarsa qualità.	**Misto** · persone fissano i criteri, filtri automatici li applicano
3	**Pre-training**	Il modello apprende relazioni statistiche nei dati, ad esempio prevedendo il token successivo.	**Autonomo** · nessun intervento umano sui singoli esempi
4	**Supervised Fine-Tuning (SFT)**	Il modello viene addestrato su esempi di istruzioni e risposte desiderate.	**Misto** · persone scrivono gli esempi, la macchina li apprende
5	**Allineamento alle preferenze**	RLHF o DPO orientano le risposte verso comportamenti più utili, corretti e sicuri.	**Misto** · persone confrontano le risposte, la macchina si adatta
6	**Valutazione e testing**	Si verificano capacità, accuratezza, robustezza e sicurezza con benchmark e test.	**Misto** · test automatici e verifiche di esperti umani
7	**Deployment e monitoraggio**	Il modello viene distribuito e monitorato per individuare problemi e miglioramenti.	**Misto** · sistemi automatici segnalano, persone analizzano e decidono

SFT Supervised Fine-Tuning, addestramento supervisionato su esempi · **RLHF** Reinforcement Learning from Human Feedback, apprendimento per rinforzo dal feedback umano · **DPO** Direct Preference Optimization, ottimizzazione diretta delle preferenze

In arancio: le fasi di addestramento vero e proprio

Che cosa sa davvero del mondo

Conosce i racconti, non le esperienze

Ha letto moltissimo sul mare, ma non ci ha mai fatto il bagno.

Si ferma a una data

Sa ciò che c'era nei suoi dati. Dopo, sa solo quello che trova cercando, se può farlo.

Vede il mondo come lo raccontano i dati

Ciò che è poco raccontato, per lei esiste poco.

Il suo mondo è una **rappresentazione**: fedele quanto lo sono i dati da cui l'ha imparato.

Bias: il modello eredita i nostri squilibri

Un modello impara dai dati. Se i dati sono sbilanciati, lo diventa anche il modello, spesso senza che nessuno se ne accorga.

Dati storici

Se in passato certi ruoli sono andati quasi solo a uomini, il modello lo tratta come una regola.

Rappresentazione

I gruppi poco presenti nei dati vengono riconosciuti e descritti peggio.

Scelte di progetto

Cosa misurare, cosa ottimizzare, su chi testare: sono decisioni umane.

Due casi documentati

Lupo o husky?

Nelle foto di addestramento i lupi erano sempre sulla neve. Il modello ha imparato a guardare lo sfondo: un husky sulla neve diventa «lupo».

Esperimento dell'Università di Washington, 2016.

Selezione del personale

Amazon abbandona un sistema sperimentale di valutazione dei CV. Addestrato su dieci anni di candidature in prevalenza maschili, penalizzava i curriculum che contenevano la parola **«women's»**.

Notizia rivelata da Reuters, 2018.

Dal farsi aiutare allo smettere di decidere

1 · Mi faccio aiutare

Penso io, l'AI suggerisce.

Chiedo idee per un tema, poi lo scrivo io.

→

2 · Faccio fare

L'AI produce, io controllo e correggo.

Una bozza di email che rileggo e sistemo.

→

3 · Delego tutto

L'AI decide, io mi limito a firmare.

Copio la risposta senza averla letta.

La domanda da farsi: **se smetto di farlo io, che cosa smetto di saper fare?**

5 · QUANDO L'AI SBAGLIA, MA SEMBRA AVERE RAGIONE

Plausibile e vero coincidono spesso. Non sempre.

Un modello calcola ciò che è probabile, sulla base dei dati da cui ha imparato. Da qui nascono i suoi rischi.

Allucinazioni: falso, ma convincente

Il modello produce un'informazione inventata con lo stesso tono sicuro di quelle vere: fonti, date, citazioni, numeri.

Perché succede

- Completa il testo in modo plausibile, senza consultare un archivio di fatti
- È addestrato a rispondere, raramente a dire «non lo so»
- Più la domanda è specifica o rara, più il rischio cresce

Mata v. Avianca, New York, 2023

Due avvocati depositano in tribunale una memoria che cita sei precedenti mai esistiti, generati da ChatGPT.

Il giudice li sanziona con **5.000 dollari**.

Fonte: Mata v. Avianca, Inc., Tribunale distrettuale federale di New York, giugno 2023

Tre domande davanti a ogni risposta

1

Da dove viene?

Posso risalire a una fonte che esiste davvero e che posso aprire?

2

Come lo verifico?

Trovo la stessa informazione in una fonte indipendente?

3

E se fosse sbagliata?

Più alta è la posta in gioco, più serve il controllo di una persona.

Bozze, idee e sintesi: sì. La decisione finale resta a noi.

Cosa cambia a casa, a scuola e nel lavoro

A casa

Assistenti vocali, consigli, ricette, domande di ogni giorno.

Attenzione: a cosa raccontiamo e a cosa lasciamo decidere.

A scuola

Spiegazioni su misura, esercizi, aiuto per chi è in difficoltà.

Attenzione: il compito svolto dall'AI non insegna nulla a chi lo consegna.

Nel lavoro

Bozze, sintesi, traduzioni: meno tempo sulle attività ripetitive.

Attenzione: la responsabilità del risultato resta nostra.

I nostri dati: dove vanno?

Con un chatbot online

- Il testo arriva sui server del fornitore
- Può essere conservato e, a seconda delle impostazioni, usato per l'addestramento
- Fuori dall'Unione Europea valgono anche altre leggi

Con un modello locale

- I dati restano sul nostro computer
- Funziona anche senza connessione
- Richiede un computer adeguato ed è in genere meno potente

Regola pratica: in un chatbot non scrivere ciò che non manderesti per email a un estraneo.

Modelli aperti: Llama, Mistral, Gemma, Qwen · Strumenti: Ollama, LM Studio

Rischi, opportunità, falsi allarmi

Rischi reali

- Disinformazione e falsi realistici (deepfake)
- Delegare troppo e smettere di pensare
- Bias e uso dei nostri dati

Opportunità

- Conoscenza più accessibile a tutti
- Supporto a chi ha difficoltà o disabilità
- Aiuto nella ricerca, nella medicina, nel lavoro

Paure da ridimensionare

- «È cosciente e ha intenzioni proprie»
- «Ci sostituirà in tutto, da domani»
- «O è infallibile, o è inutile»

La paura utile è quella che ci rende più attenti.

9 · E SE LA DOMANDA PIÙ IMPORTANTE RIGUARDASSE NOI?

Che cosa vogliamo continuare a saper fare, decidere e scegliere da soli?

L'AI cambierà molte cose. Quale umanità costruire insieme a lei resta una scelta nostra.

Che cosa abbiamo imparato

INTRO

Cos'è l'AI generativa

Prevede la parola più probabile: non è una mente

1

Ha sempre una risposta

Chiedere è più comodo, ma il confronto tocca a noi

2

Perché ci fidiamo

Parla bene e non esita mai: la fiducia va calibrata

3

Conosce il mondo?

Conosce i suoi dati, con i loro squilibri

4

Chiedere o delegare

Farsi aiutare sì, smettere di decidere no

5

Sbaglia e convince

Plausibile non vuol dire vero

6

Senza smettere di pensare

Da dove viene? Come lo verifico? E se fosse sbagliata?

7

Vita quotidiana

Utile ovunque, ma ogni domanda è un dato

8

Avere paura?

Attenzione ai rischi reali, non alle paure da film

9

E noi?

Le scelte che contano restano nostre

L’AI ha sempre una risposta. Le domande giuste restano nostre.

Domande?

Rimaniamo in contatto

Gianluca Garofalo

SITO

ggarofalo-dev.github.io

EMAIL

Gianluca.Garofalo@gmail.com

Inquadra il codice per aprire il sito