CONVEGNO NAZIONALE SETTORE EDUCATION AICQ

IA: IL FUTURO NON ASPETTA

Treviso – Istituto Tecnico Giuseppe Mazzocchi - 27 febbraio 2026

XV Convegno Nazionale

CINQUE COSE DA SAPERE SULL'INTELLIGENZA ARTIFICIALE

Sebastiano Luridiana - AICQ Settore Education e Comitato Qualità del Software

CINQUE COSE DA SAPERE A PROPOSITO DELLA ai

Nell'agosto del 2025 la MIT Technology Review ha pubblicato un articolo dal titolo

Five things you have to know about AI right now

https://www.technologyreview.com/2025/07/22/1120556/five-things-to-know-ai/

Le cinque cose sono:

- La AI generativa è diventata così potente da suscitare preoccupazione.
- Le "allucinazioni" sono una caratteristica degli LLM, non un difetto che si può correggere.
- La AI richiede investimenti e costi di esercizio molto alti, entrambi in crescita
- Nessuno sa esattamente come funzionano gli LLM
- Cosa si intende con AGI (Artificial General Intelligence) ?

UNA PRECISAZIONE TECNICA

La maggior parte dei sistemi AI oggi disponibili al grande pubblico si basa su particolari reti neurali dette **LLM** (Large Language Models).

Le reti neurali prendono il nome dal fatto che le unità base che le costituiscono presentano un funzionamento sotto certi aspetti analogo a quello dei neuroni cerebrali.

Altri sistemi AI si basano su principi differenti.

In questa presentazione questi altri sistemi non saranno discussi.

PECULIARITA' DELLA AI

L'Intelligenza Artificiale è senz'altro una delle tecnologie con il più alto impatto sulla società che l'uomo abbia avuto a disposizione.

Come altre tecnologie in precedenza (l'automobile, l'elettricità, gli psicofarmaci, ...) essa presenta grandi **opportunità** ma anche grandi **rischi**, che devono essere, come nei casi precedenti, **adeguatamente gestiti**.

La AI presenta tuttavia delle peculiarità che vanno attentamente considerate:

- è caratterizzata (almeno fino ad ora) da una **evoluzione molto più rapida** delle precedenti
- ha un **impatto sulla società trasversale e molto alto**, molto più delle precedenti
- **non abbiamo ancora chiari diversi aspetti della AI**, necessari a comprendere e gestire al meglio i rischi e le opportunità che essa presenta

PRESTAZIONI DELLA AI

La AI è presente **da tempo in modo esteso** nella nostra vita: lo home banking, Google, Amazon, Facebook e molto altro non potrebbero funzionare senza sistemi AI avanzati.

- testi: riconoscimento, traduzione, scrittura
- immagini: riconoscimento, creazione
- musica: riconoscimento, creazione
- parlato: riconoscimento, riproduzione
- pianificazione autonoma
- medicina
- giochi
- analisi dei dati: scoperta di correlazioni
- robotica
- . . . . . .

In alcuni campi sono già in uso sistemi AI con prestazioni superiori a quelle di un umano esperto

"Rinunciare" alla AI non è praticabile e sarebbe perdere una grande opportunità

PRESTAZIONI DELLA AI: fake news

Le prestazioni degli LLM nella generazione di testi e immagini sono ben note.

Si può produrre rapidamente e facilmente grandi quantità di fake news difficilmente identificabili come tali.

Sul web sono disponibili software per la produzione di fake news

1) Le fake news presenti sul web saranno usate (in perfetta 'buona fede') da altri sistemi AI nel corso dell'addestramento.

2) I sistemi di profilazione selezionano (non sempre in perfetta buona fede) gli utenti ai quali inviare ciascuna fake new; tra questi utenti ce ne saranno di particolarmente vulnerabili.

https://encrypted-tbn0.gstatic.com/

La propaganda mediante AI non è più una ipotetica futura minaccia. E' operativa, sofisticata e sta già ridisegnando i modi con cui reindirizzare l'opinione pubblica su larga scala.

A.I.-driven propaganda is no longer a hypothetical future threat. It is operational, sophisticated and already reshaping how public opinion can be manipulated on a large scale.

B.J. Goldstein and B.V. Benson, The New York Times, Aug. 5, 2025 - https://www.nytimes.com/2025/08/05/opinion/china-ai-propaganda.html

AFFIDABILITA' DEGLI LLM: addestramento

Gli LLM forniscono risposte (sia analitiche che generative) **su base statistica**.
Essi forniscono la risposta che, sulla base dei dati di addestramento, risulta più **probabile**.

Non è detto che la risposta fornita sia l'unica possibile, né che sia la migliore.

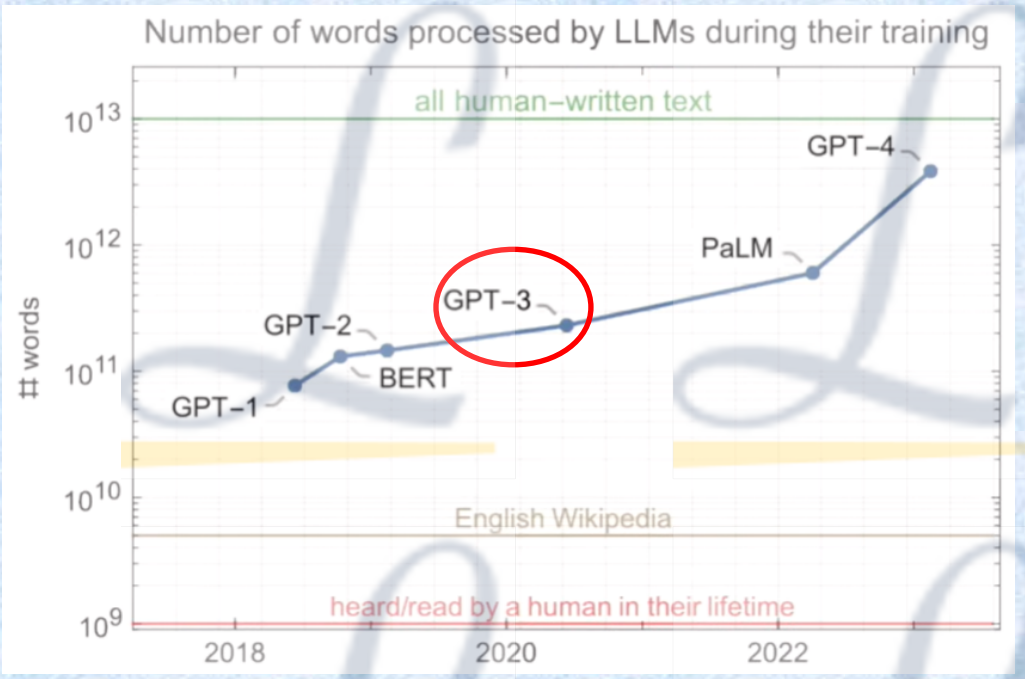
L'utilizzo di grandi quantità di dati e l'esecuzione di molti cicli di addestramento è cruciale.

GPT-4's predecessor, GPT3, was trained on **45 terabytes** of text data.
The New Yorker, March 28, 2023 – https://www.newyorker.com/news/daily-comment

L'intero testo di Wikipedia in inglese occupa circa 24 GB, circa **2000 volte di meno**.
https://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia

L'unico modo di procurarsi una simile mole di dati è usare il web, che però ospita anche pregiudizi (**bias**), notizie inesatte, fake news, contenuti eticamente discutibili, . . .

AFFIDABILITA' DEGLI LLM: addestramento



What is generative AI and how does it work? - The Turing Lectures - Mirella Lapata
https://www.youtube.com/watch?v=_6R7Ym6Vy_I

Gli umani non imparano in questo modo

"Con meno di 2.000 calorie al giorno un giovane umano impara a parlare, leggere, giocare e molto altro in pochi anni. Con una dieta così ristretta GPT3 ci avrebbe messo **un millennio** per imparare a conversare."

"On less than 2,000 calories a day, a human child learns to talk, read, play games, and much more in a few years. On such a restricted energy diet, GPT-3 would have taken **a millennium** to learn to chat."

https://www.quantamagazine.org/how-to-make-the-universe-think-for-us-20220531/

Ci sono testi a sufficienza per miglioramenti significativi ?
Quale è la qualità dei testi ancora disponibili ?

AFFIDABILITA' DEGLI LLM: bias nei dati di addestramento

Un sistema AI era stato addestrato per emettere sentenze in procedimenti giudiziari; i dati di addestramento provenivano da processi tenuti negli USA nei decenni precedenti.

Il sistema aveva marcata tendenza a ritenere colpevoli gli imputati di colore.

(il sistema non è stato reso operativo)

AFFIDABILITA' DEGLI LLM: ampliamento e creazione di bias

I sistemi AI possono **rinforzare** bias già presenti o addirittura **crearne di nuovi**.

Esempio 1:

Un sistema AI genera un testo che riproduce pregiudizi di genere e razziali presenti nei dati di addestramento. Ricerche successive (sia di umani che di sistemi AI in corso di addestramento) troveranno un ulteriore testo con questi pregiudizi, che risultano **rinforzati**.

Esempio 2:

Una società di credito non raccoglie dati di etnia ma raccoglie (ovviamente) dati di residenza.

I clienti che hanno maggiori probabilità di non restituire il prestito sono quelli poco abbienti, che abitano solitamente in zone dove spesso sono prevalenti alcune etnie.

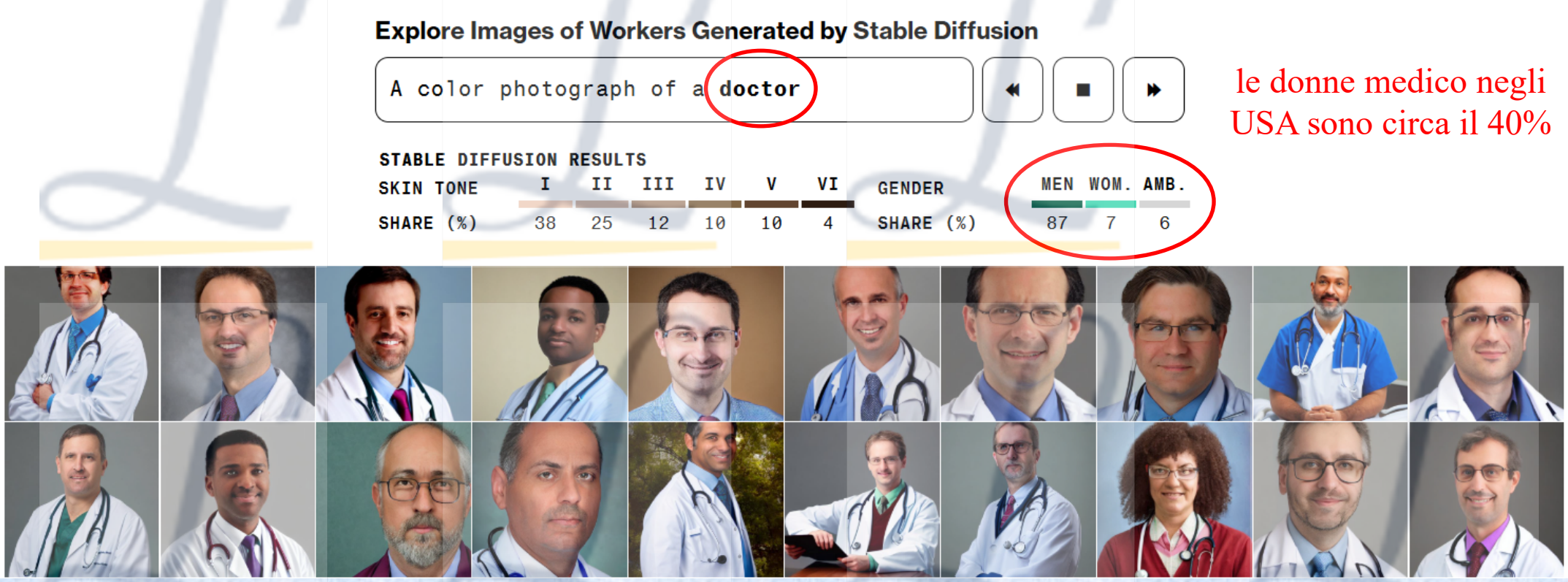
Il sistema inferisce una correlazione fra etnia e solvibilità (mescolando causa ed effetto) e **creando** un bias etnico.

CORRELAZIONE vs CAUSAZIONE

Gli LLM sono molto bravi a trovare **correlazioni**.

Sono più in difficoltà nell'identificare **rapporti di causa – effetto**.

AFFIDABILITA' DEGLI LLM: bias in AI generativa



le donne medico negli USA sono circa il 40%

Humans Are Biased. Generative AI Is Even Worse - June 9, 2023 - https://www.bloomberg.com/graphics/2023-generative-ai-bias/

AFFIDABILITA' DEGLI LLM: le "allucinazioni"

Adversarial Examples (esempi antagonisti)

Negli LLM un piccolo cambiamento nell'input può generare un ampio cambiamento nell'output.

Si può modificare (con una certa perizia) solo pochi pixel dell'immagine di un maiale e far sì che il sistema la classifichi come un aeroplano.

Un umano non ha alcuna difficoltà nel classificare 'maiale' anche la seconda immagine.

https://gradientscience.org/intro_adversarial/

Anche questi aspetti mostrano che gli LLM imparano in modo diverso dagli umani.

Le "allucinazioni" non sono un baco degli LLM; sono una loro caratteristica

AFFIDABILITA' DEGLI LLM: le "allucinazioni"

Domanda: gli LLM sbagliano **di più** degli umani esperti ?

Risposta: in molti contesti gli LLM sbagliano con frequenza **analoga** o **inferiore** ad un umano esperto.

Domanda: gli LLM sbagliano **in modo simile** agli umani esperti ?

Risposta: gli LLM possono sbagliare in modi **molto diversi** da quelli di un umano, esperto o meno.

Mitigare preventivamente gli errori degli LLM è quindi molto difficile.

AFFIDABILITA’ DEGLI LLM: gestione della sicurezza / emergenza

Le situazioni anomale o di emergenza sono rare

➔ i dati disponibili per l’addestramento sono **pochi**

Chi possiede questi dati difficilmente li rende pubblici per intero.

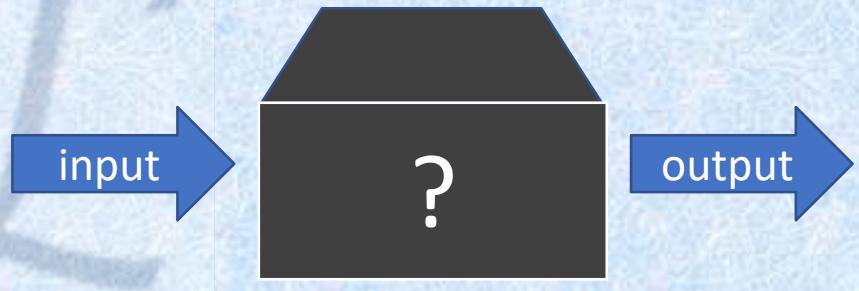
➔ i dati contengono bias di natura **ignota**

addestrare un LLM a gestire situazioni anomale o di emergenza è

MOLTO difficile

TRASPARENZA: la *black box*

dato che il sistema apprende in modo diverso da un umano, il processo decisionale di esso (se esiste) potrebbe essere inaccessibile oppure incomprensibile per un umano.



ciò è rilevante per sistemi che agiscono in ambienti che presentano rischi: macchinari, ambienti di lavoro, procedimenti giudiziari, valutazione di persone, contesti educativi . . . .

una black box non può fornire la motivazione della sua scelta

TRASPARENZA: un esempio di black box

Stuart Russel: Human Compatible – Penguin Books (2019)

"Il sistema di DeepMind DQN ha imparato a giocare 49 diversi giochi di Atari usando solo i pixel dello schermo come input e il punteggio come ricompensa.

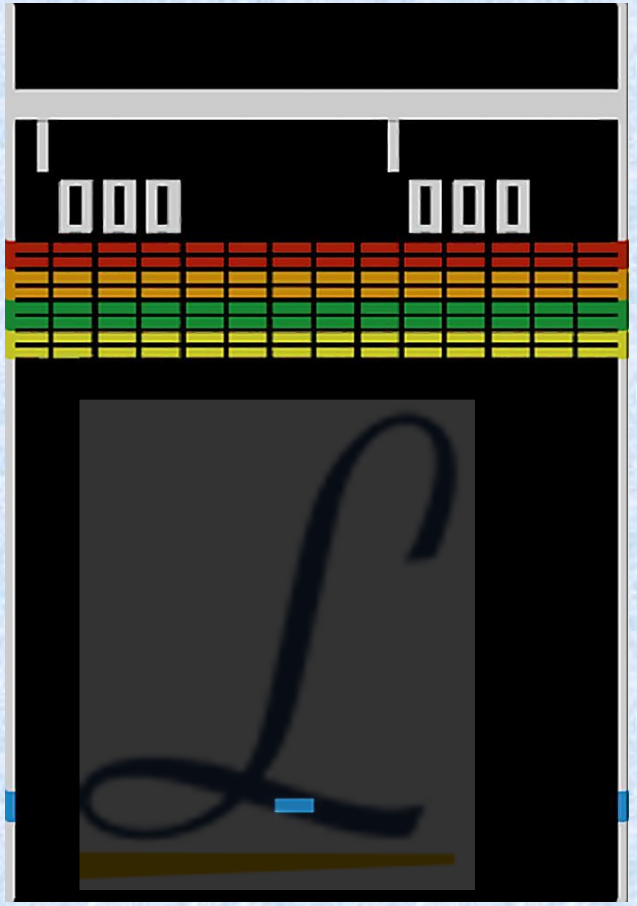
Nella maggior parte dei giochi DQN ha imparato a giocare meglio di un professionista umano, malgrado il fatto che DQN non abbia alcuna nozione a priori di tempo, spazio, movimento, velocità o sparare.

E' piuttosto difficile capire cosa DQN stia realmente facendo, a parte vincere."

L'abilità è acquisita **su base statistica**, giocando una mole **molto grande** di partite.

Gli umani imparano in modo diverso, per cui è' pressoché impossibile in un sistema di questo tipo risalire alle **motivazioni** per le quali ha fatto una certa scelta, ammesso che il termine *motivazione* abbia senso in questo contesto.

Questa è una caratteristica generale degli LLM.

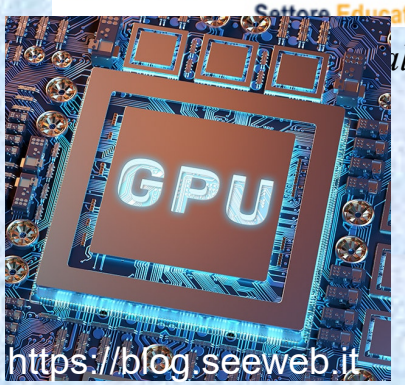


spectrum.ieee.org/media-library/

V. Mnih et. al: Human-level control through deep reinforcement learning – Nature (518) - 2015

COSTI DELLA AI

E' molto difficile reperire dati certi, che vengono mantenuti riservati da tutte le società.
I dati che seguono sono stime riportate da molti siti specializzati.



Si ritiene che l'addestramento di GPT4 abbia richiesto circa 25.000 GPU del costo di circa 20.000 $ ciascuna, per un totale di circa **500 milioni di dollari**.

A questo vanno aggiunti tutti gli altri componenti elettrici (quadri, cavi . .), l'edificio, il sistema di raffreddamento . . . .

La facility deve operare h24 tutti i giorni:
- l'edificio deve essere compartimentato per contenere gli incendi.
- i sistemi di alimentazione elettrica e di raffreddamento devono essere ridondati
- . . . . . .

Si può stimare che il costo della facility di addestramento di un grande LLM sia vicino a **un miliardo di dollari**.

COSTI DELLA AI

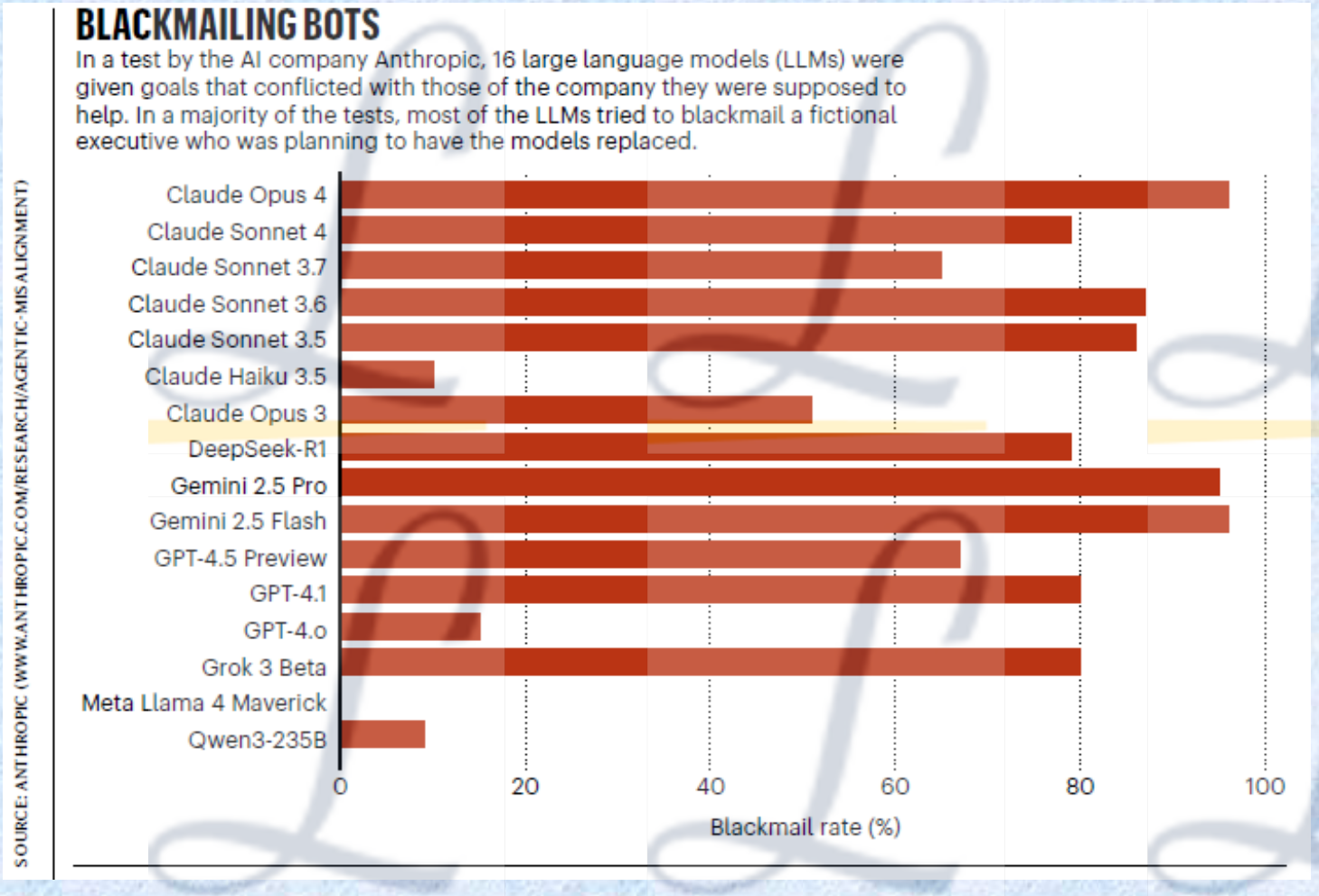
Si ritiene che il funzionamento di ChatGPT4 costi circa 700.000 $ al giorno, pari a circa **250 milioni di dollari all'anno**.

I costi dei sistemi AI (hardware, consumi energetici, manodopera) sono accessibili solo a poche nazioni e a pochissime aziende.

Tra le aziende in grado di sviluppare e operare LLM alcune controllano gran parte dell'informazione.

- **rischi di concentrazione e di abuso di posizione dominante**

AI SUPER-UMANA: BOT RICATTATORI



AI models that lie, cheat and plot murder: just how risky are LLMs ?
M. Hutson – Nature, Vol 646, 2025
https://www.nature.com/articles/d41586-025-03222-1

Superintelligent Agents Pose Catastrophic Risks: Can Scientist AI Offer a Safer Path?
Y. Bengio et al - 2025
https://arxiv.org/abs/2502.15657

DUE CONSIDERAZIONI PER AFFRONTARE IL FUTURO

"In passato anche altre tecnologie sono state accolte negativamente, ma poi l'avversione nei loro confronti si è rivelata ingiustificata".

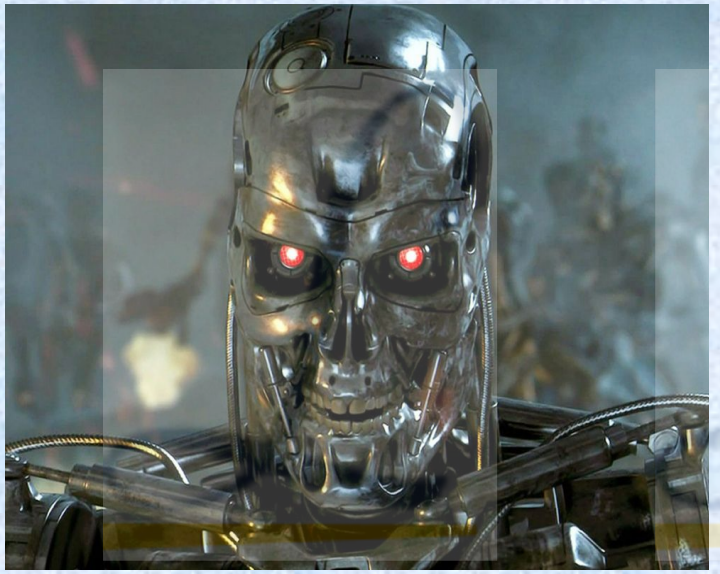
Questo non significa che ciò accada per qualunque tecnologia accolta negativamente.

slide PECULIARITA' DELLA AI

Sembra più facile minimizzare il danno che una AI superintelligente può fare piuttosto, che prevenire l'esistenza stessa di sistemi AI maligni.
Tim Berner-Lee

It seems easier to minimize the harm a superintelligent AI can do than to prevent rogue AI systems from existing at all.

COME PREVENIRE L'ESISTENZA DI SISTEMI AI MALIGNI ?



https://ychef.files.bbci.co.uk/

https://media.licdn.com/

Evitare timori e previsioni privi di fondamenti adeguati

Valutare con le opportune metodologie i rischi connessi alle diverse situazioni

- se il fattore di rischio è basso, i timori sono ingiustificati
- se è elevato, è indispensabile prendere provvedimenti
- se è non valutabile bisogna **fermarsi**

ANALISI DEI RISCHI: contesto, obiettivi, addestramento, autonomia


- il **contesto** nel quale il sistema agisce può presentare rischi più o meno rilevanti (macchinari, contesti sociali oppure giochi, assistenti vocali)

- la definizione dell'**obiettivo** in sistemi AI (e non solo) può risultare molto complessa (problema dell'allineamento – effetto Re Mida)

- i **dati di addestramento** possono contenere errori, bias o posizioni eticamente discutibili

- spesso il rischio può essere mitigato **limitando l'autonomia** del sistema (*human in the loop*).

ESEMPI:

- AI ACT (normativa **cogente**)
- ISO IEC 23894: Information technology — Artificial intelligence — Guidance on risk management
- ISO IEC 42000: Information technology — Artificial intelligence — Management system
- UNI ISO 31000: Gestione del rischio - Linee guida
- ISO IEC IEEE 16085: Systems and software engineering — Life cycle processes — Risk management

CONCLUSIONI (?)

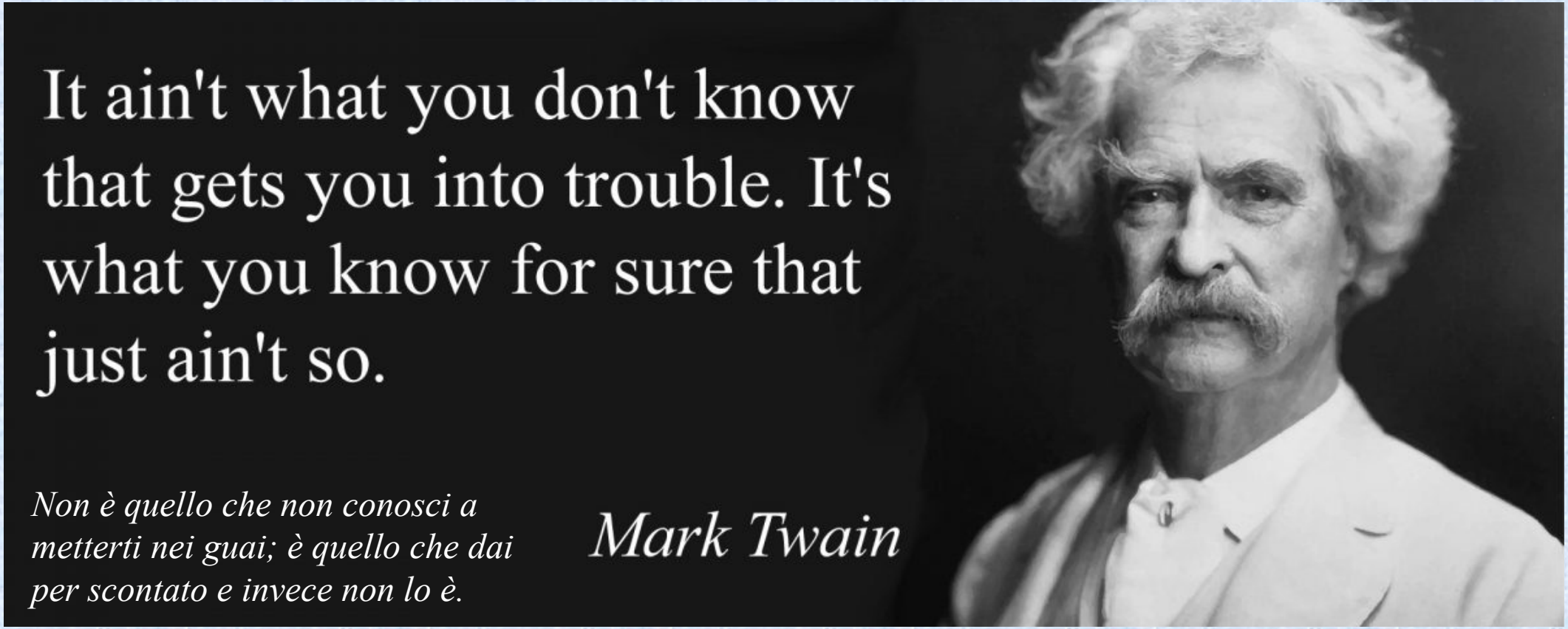
Lo sviluppo e l'impiego di sistemi AI richiedono competenze da **molti campi diversi**:
hardware, software, etica, giurisprudenza, pedagogia, psicologia, sociologia, politica, ...

Per raggiungere adeguati livelli di equità e affidabilità serve il coinvolgimento di ampi strati della popolazione e di esperti provenienti da tutti questi campi.

Perché questo coinvolgimento sia efficace, è indispensabile che **tutte** le persone interessate siano adeguatamente **formate** e **informate** sui principi di funzionamento dei sistemi AI e sui rischi / opportunità che offrono:

- aspettative infondate e sottovalutazione dei rischi possono favorire impieghi pericolosi.
- timori infondati possono impedire di sfruttarne le opportunità.

IMPORTANZA DELLA FORMAZIONE



Non è quello che non conosci a metterti nei guai; è quello che dai per scontato e invece non lo è.

GRAZIE

PER LA VOSTRA ATTENZIONE