
Intelligenza Artificiale

Ma di cosa stiamo parlando?

Che cos'è l'*Intelligenza Artificiale*?

L'intelligenza artificiale nel suo significato più ampio, è la capacità o il tentativo di un sistema artificiale (tipicamente un sistema informatico) di simulare una generica forma di intelligenza.

- if temperatura > 24 and sole → abbassa tapparelle.
- hai comprato questo libro, potrebbe piacerti anche...
- riconoscimento targhe nei parcheggi
- “Aqua vite” → Intendevi forse “acqua vite”?
- ...

Che cos'è l'*Intelligenza Artificiale* **Generale**?

L'intelligenza artificiale “forte” o intelligenza artificiale generale (in sigla AGI dall'inglese artificial general intelligence) è la capacità di un agente intelligente di apprendere e capire un qualsiasi compito intellettuale che può imparare un essere umano.

- Non esiste ancora
- Esiste il famoso test di Turing
- ... qualcuno inizia a dire che ci siamo vicini

Breve storia dell'Intelligenza Artificiale

1. Origini e concetti precursori

- **Antichità e mitologia:** L'idea di macchine "intelligenti" appare in miti come il gigante Talos della mitologia greca e automi progettati da inventori come Erone di Alessandria.
- **Filosofia:** Concetti come la logica formale (Aristotele) e la meccanizzazione del pensiero (Cartesio e Leibniz) gettano le basi teoriche.

Breve storia dell'Intelligenza Artificiale

2. Nascita dell'intelligenza artificiale moderna (anni '40-'50)

- **Macchine calcolatrici:** Lo sviluppo di computer programmabili permette l'elaborazione automatica.
- **Alan Turing:** Introduce il "Test di Turing" (1950) per valutare l'intelligenza di una macchina e formalizza il concetto di "macchina universale".
- **Conferenza di Dartmouth (1956):** John McCarthy conia il termine "intelligenza artificiale". Questo evento segna l'inizio ufficiale della disciplina.

Breve storia dell'Intelligenza Artificiale

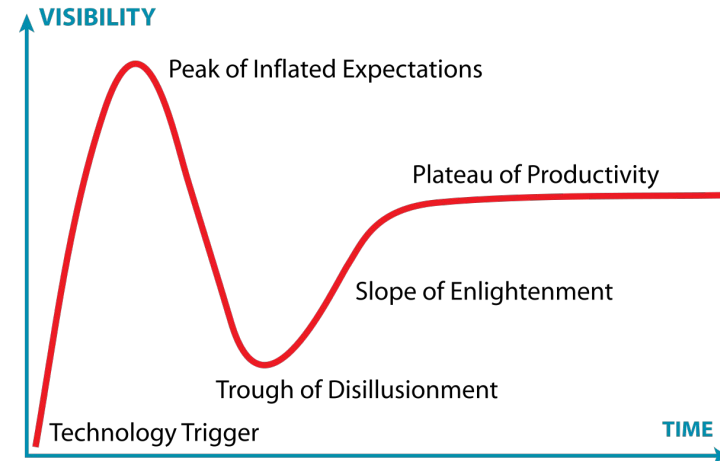
3. I primi successi (anni '50-'70)

- **Logica e giochi:**
 - Primo programma di scacchi (Claude Shannon e Alan Turing).
 - Logic Theorist (1956) di Allen Newell e Herbert Simon risolve teoremi matematici.
- **Linguaggio naturale:** Il programma ELIZA (Joseph Weizenbaum, 1966) simula una conversazione con un terapeuta.
- **Sistemi esperti:** Primi tentativi di creare sistemi basati su regole per risolvere problemi specifici.

Breve storia dell'Intelligenza Artificiale

4. L'inverno dell'IA (anni '70-'80)

- **Limitazioni tecnologiche:** Mancanza di potenza computazionale e difficoltà nella gestione dei dati complessi rallentano i progressi.
- **Aspettative irrealistiche:** Promesse non mantenute portano a tagli nei finanziamenti e disillusione.

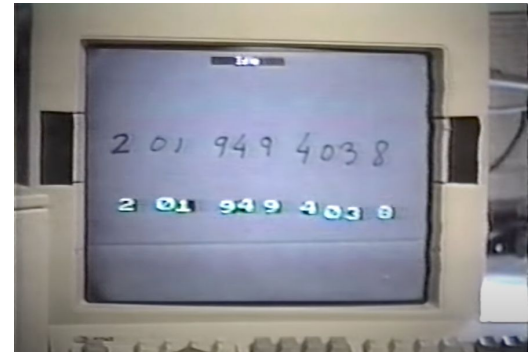


Breve storia dell'Intelligenza Artificiale

5. Rinascita e crescita (anni '80-'90)

- **Sistemi esperti:** l'innovazione hardware porta ai primi successi software come ad esempio MYCIN (diagnosi mediche), mostrando applicazioni pratiche concrete.
- **Reti neurali:** Riscoperta degli algoritmi di apprendimento come il backpropagation (1986) rinnova l'interesse per le reti neurali.
- **Robotica:** Sviluppi significativi nell'automazione e nei robot industriali.

https://www.youtube.com/watch?v=FwFduRA_L6Q



Breve storia dell'Intelligenza Artificiale

6. L'era del machine learning (anni 2000)

- **Big Data:** La disponibilità di enormi quantità di dati e l'aumento della potenza computazionale (GPU) rendono praticabili algoritmi complessi.
- **Apprendimento automatico:** tecniche come support vector machines (SVM) e clustering migliorano le capacità analitiche.
- **Applicazioni pratiche:** Applicazioni in motori di ricerca, sistemi di raccomandazione (Amazon, Netflix), e primi assistenti virtuali su compiti specifici (Siri, 2011, Alexa).

Breve storia dell'Intelligenza Artificiale

7. L'era delle reti neurali profonde (Deep Learning) (anni 2010)

- **Avanzamenti tecnici:**
 - Algoritmi come il deep learning (reti neurali profonde) portano a progressi in riconoscimento vocale, visione artificiale e traduzione automatica.
 - Tecnologie come GPT e BERT rivoluzionano il trattamento del linguaggio naturale
- **Applicazioni non banali:** in campi come la diagnostica medica, riconoscimento facciale e i veicoli autonomi.
- **Eventi simbolici:**
 - AlphaGo (DeepMind, 2016) batte il campione mondiale di Go e nascita di **OpenAI** (2015)

Breve storia dell'Intelligenza Artificiale

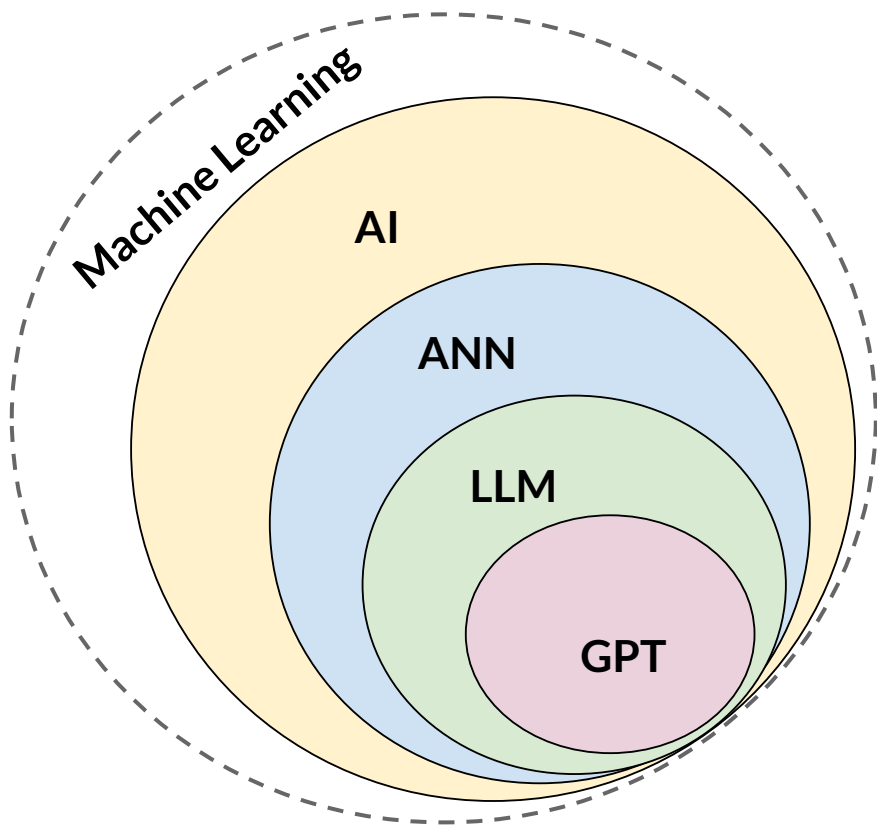
8. L'IA generativa e l'automazione avanzata (anni 2020 e oltre)

- **Modelli generativi multimodali:** Sistemi come GPT-4 e DALL-E combinano immagini, testo e altre forme di dati e danno una *percezione* di intelligenza artificiale avanzata, a tratti generale.
- **Etica e regolamentazione:**
 - Cresce l'interesse per temi etici, trasparenza e impatti sociali dell'IA.
 - Legislazioni come l'AI Act dell'Unione Europea cercano di regolare l'uso dell'IA.
- **Prospettive future:**
 - IA generale (AGI): Ricerca verso sistemi con intelligenza simile a quella umana.
 - Collaborazione uomo-macchina: Integrazione sempre più profonda tra IA e attività umane.

Breve storia dell'Intelligenza Artificiale

“Tutto quello che ci hanno sempre detto non avrebbe mai funzionando sta iniziando a funzionare”

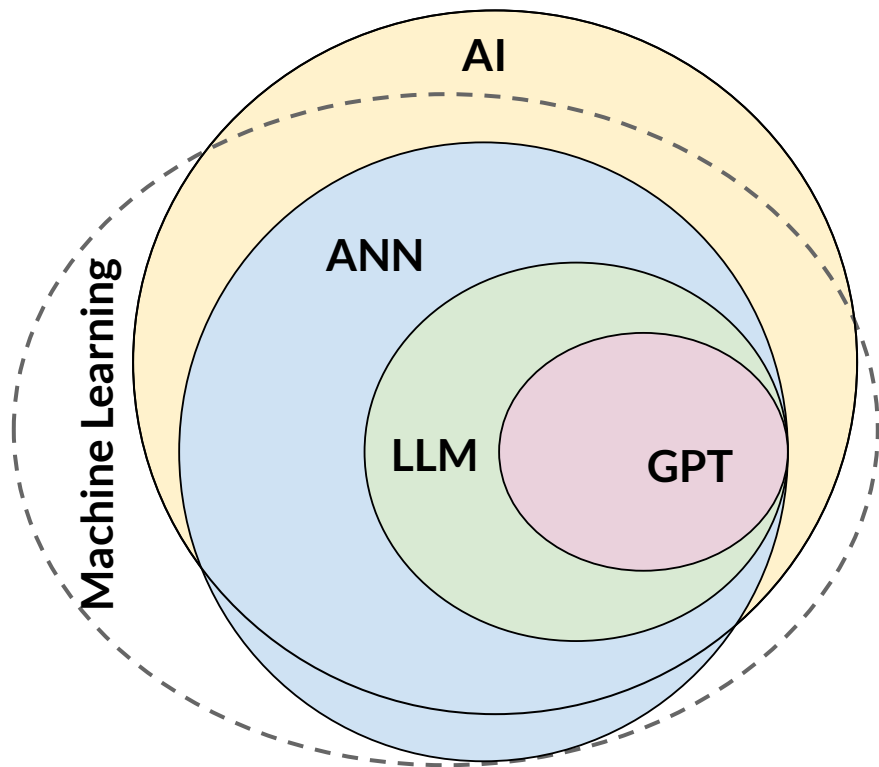
Machine learning, AI, Reti Neurali, LLM, GPT?



- Artificial Intelligence
- Artificial Neural Networks
- Large Language Models
- Generative Pre-trained Transformers

(circa, se il mondo fosse semplice)

Machine learning, AI, Reti Neurali, LLM, GPT?



- Artificial Intelligence

- Artificial Neural Networks

- Large Language Models

- Generative Pre-trained Transformers

(com'è veramente, ma sempre circa)

Cos'è un modello

- “Con modello fisico si intende una rappresentazione concettuale (spesso una semplificazione che ammette una formalizzazione matematica) del mondo reale o di una sua parte, capace di spiegare un determinato fenomeno.”
- “Un modello matematico è una rappresentazione quantitativa di un fenomeno naturale.”
- “Un modello statistico è un modello matematico che incarna un insieme di ipotesi statistiche riguardanti la generazione di dati campione.”

→ *E' una generica ipotesi di come credete che si comporti il mondo reale*

Cosa me ne faccio di un modello?

Di base, i modelli si usano per fare ***predizioni***.

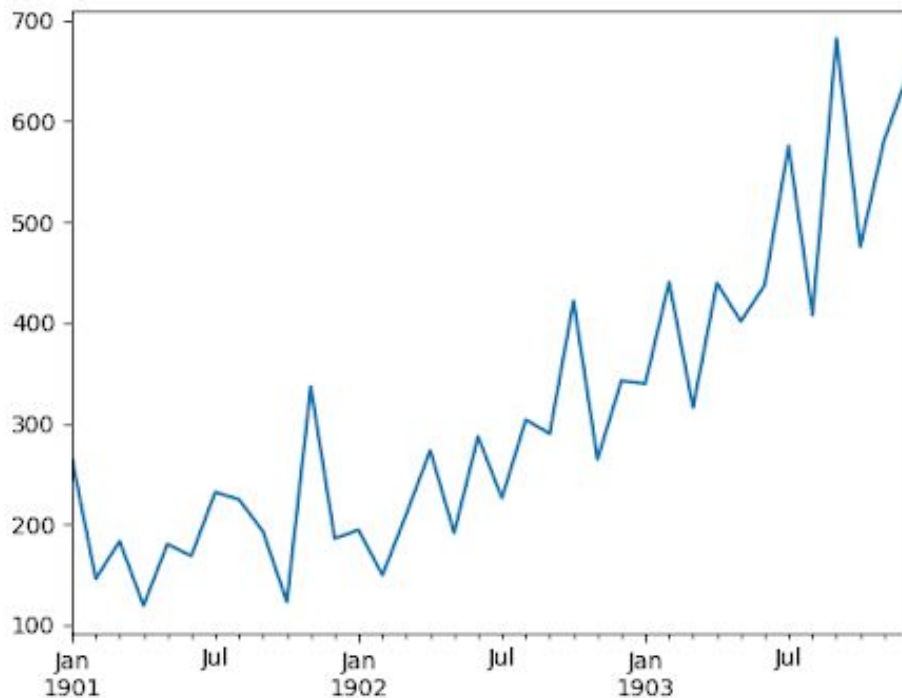
→ Come saranno le vendite il prossimo mese?

→ Quale sarà la parola che viene dopo “ciao come”?

→ Come si propagherà un’epidemia?

Esempio: le vendite dello shampoo

shampoo_sales.csv

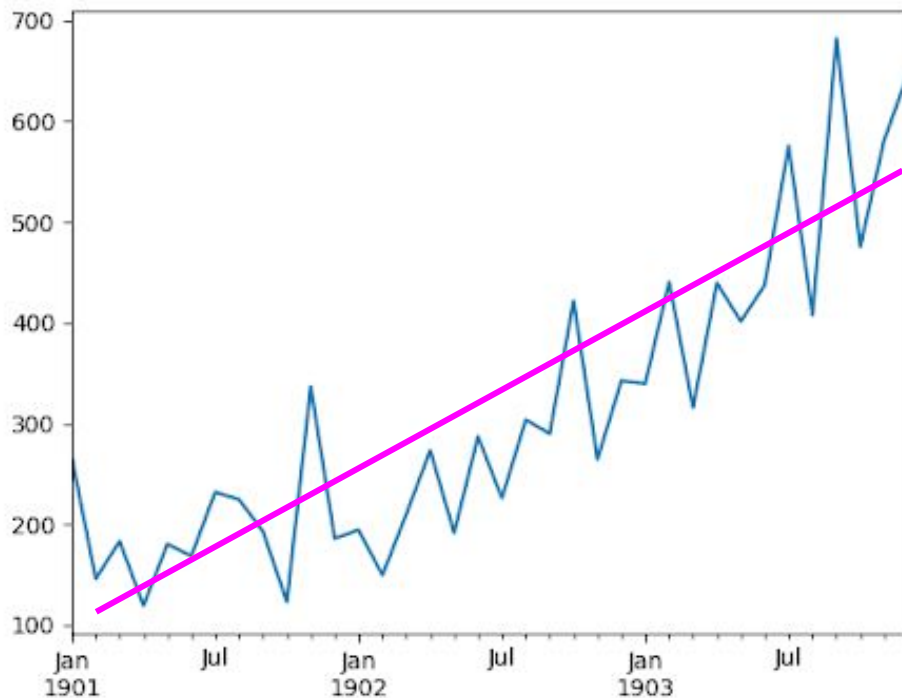


Perchè mi servono i dati per un modello?

- Posso creare anche un modello senza dati.
- Per esempio, so che le vendite di regali hanno un'impennata sotto le feste. Di quanto aumentano non lo so, ma decido a tavolino che l'aumento è del 50%. Questa è la mia ipotesi del modello.
- ..se però ho dei dati delle vendite, allora posso calcolare quanto è l'aumento rispetto al mese precedente direttamente dai dati, e magari trovare che non aumentano del 50%, ma del 67%. O del 78%.
- In tal caso si dice che si fa il **fit** del modello (sui dati).

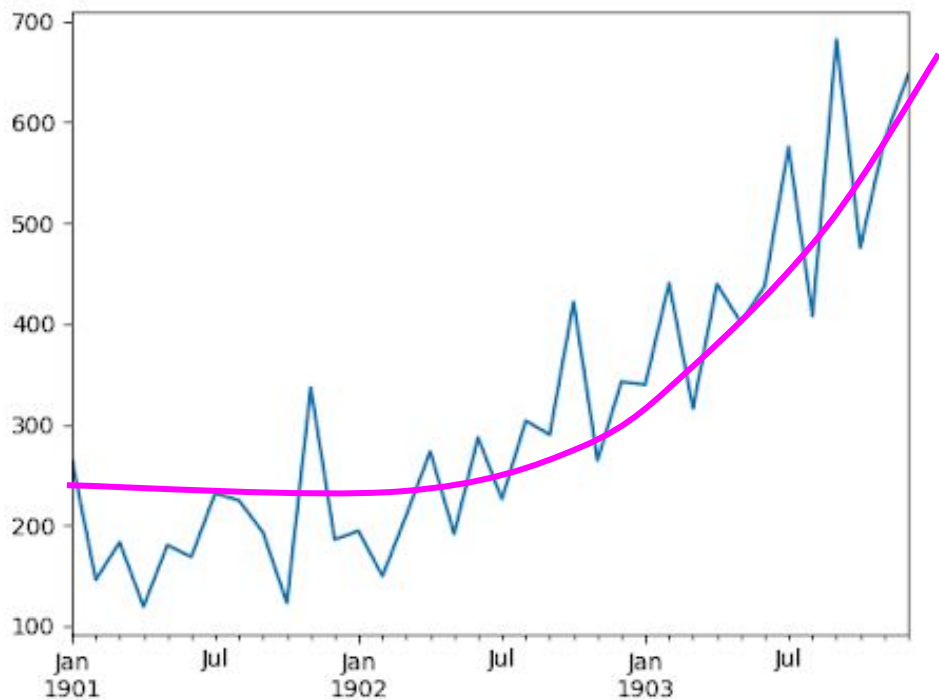
Esempio: le vendite dello shampoo (fit lineare)

shampoo_sales.csv



Esempio: le vendite dello shampoo (fit esponenziale)

shampoo_sales.csv



Le 4 macro-classi di modelli

- Deterministici (numerici/esperti)
- Statistici
- Machine learning
- Deep Learning

Deterministici (numerici/esperti)

- Modelli basati su equazioni fisiche o matematiche (numerici)
- Modelli basati su knowledge-based reasoning (esperti, regole)
- Output deterministico a partire dai dati di input

Deterministici (numerici/esperti)

- GLM
- ARIMA
- modelli bayesiani

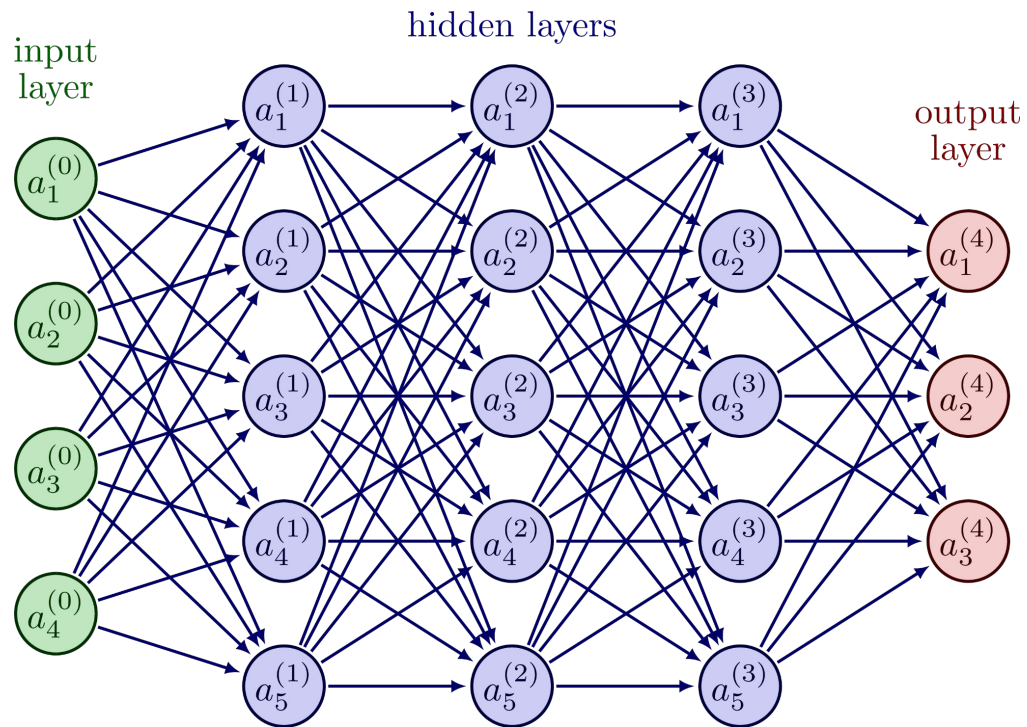
Deterministici (numerici/esperti)

Deterministici (numerici/esperti)

Le reti neurali

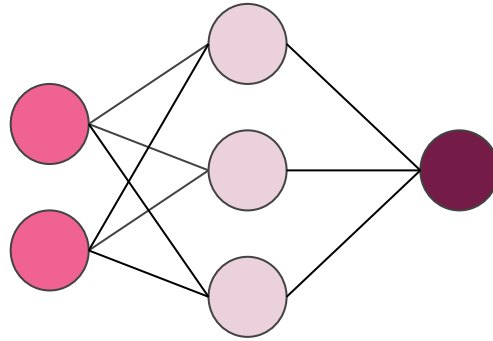
- Modello computazionale basato sulla biologia
- Simulano le connessioni neurali del cervello
- Hanno un livello di ***input***, dei livelli nascosti, ed un livello di ***output***
- I collegamenti tra i neuroni hanno dei “pesi” che permettono di mappare determinati input su determinati output

Le reti neurali



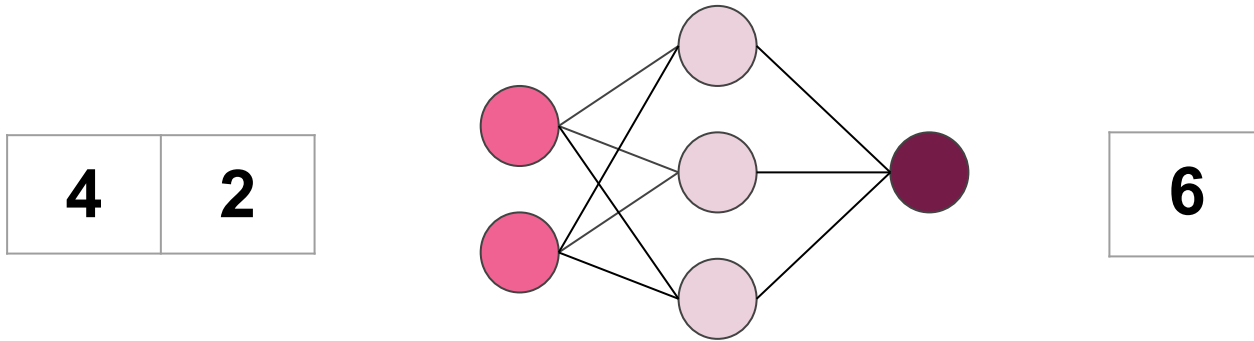
Le reti neurali: apprendimento

1	1
1	2
2	2
2	3

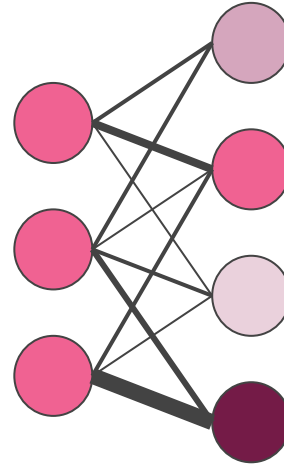
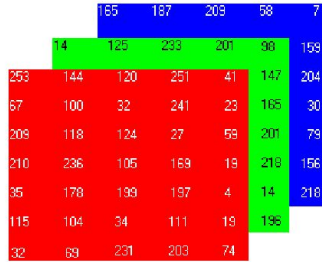


2
3
4
5

Le reti neurali: utilizzo

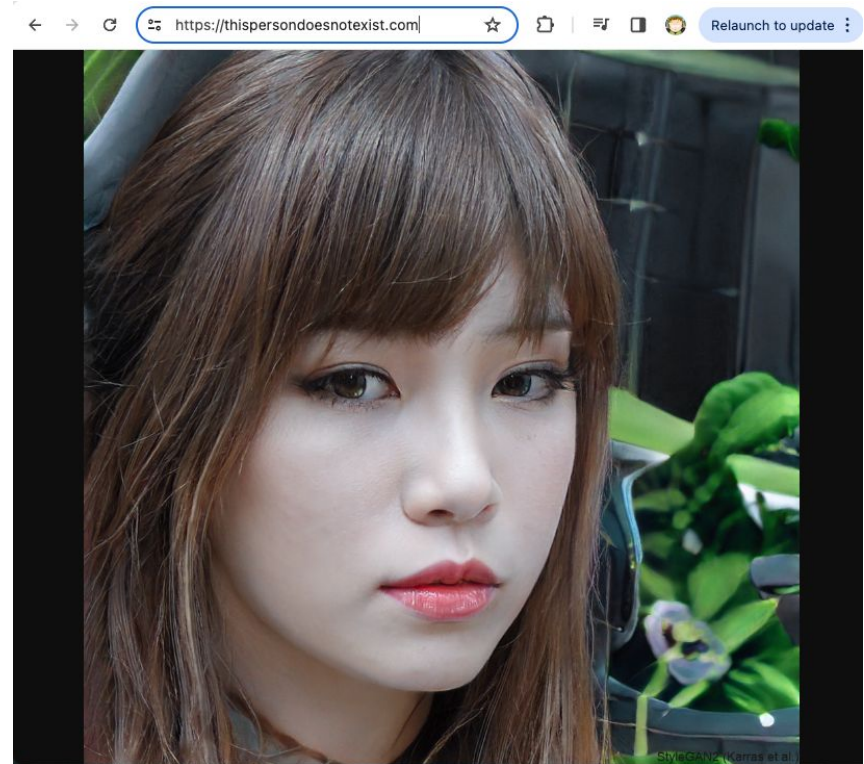


Le reti neurali e le immagini

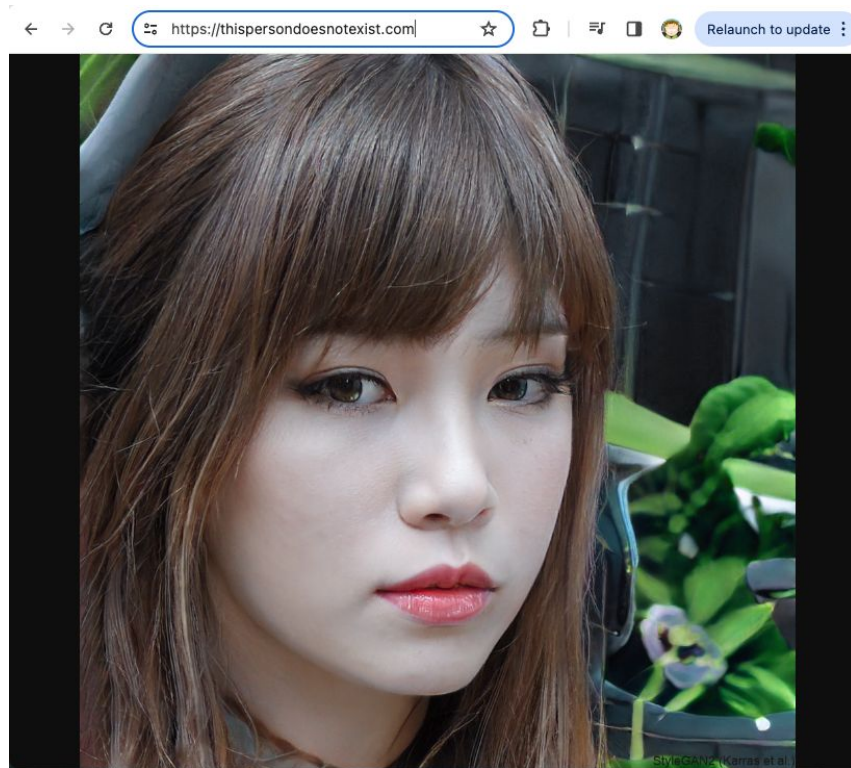
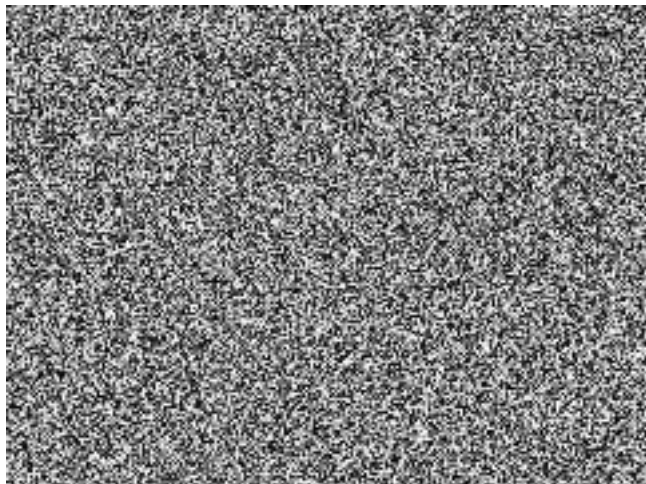


Automobile

Intelligenza Artificiale *generativa*: immagini



Intelligenza Artificiale *generativa*: immagini



Intelligenza Artificiale *generativa*: testo



Input: “Era una notte ...” prediction: “**buia**”

Input: “Era una notte **buia**” prediction: “**e**”

Input: “Era una notte **buia e**” prediction: “**tempestosa**”

oppure

Input: “Era una notte ...” prediction: “**afosa**”

Input: “Era una notte **afosa**” prediction: “**a**”

Input: “Era una notte **afosa a**” prediction: “**Milano**”

...next-token prediction

Intelligenza Artificiale *generativa*: testo

Le parole non sono numeri o pixel!

Cosa **entra** dentro una rete neurale che lavora col testo?

“Puoi farmi per punti la storia dell'intelligenza artificiale?”

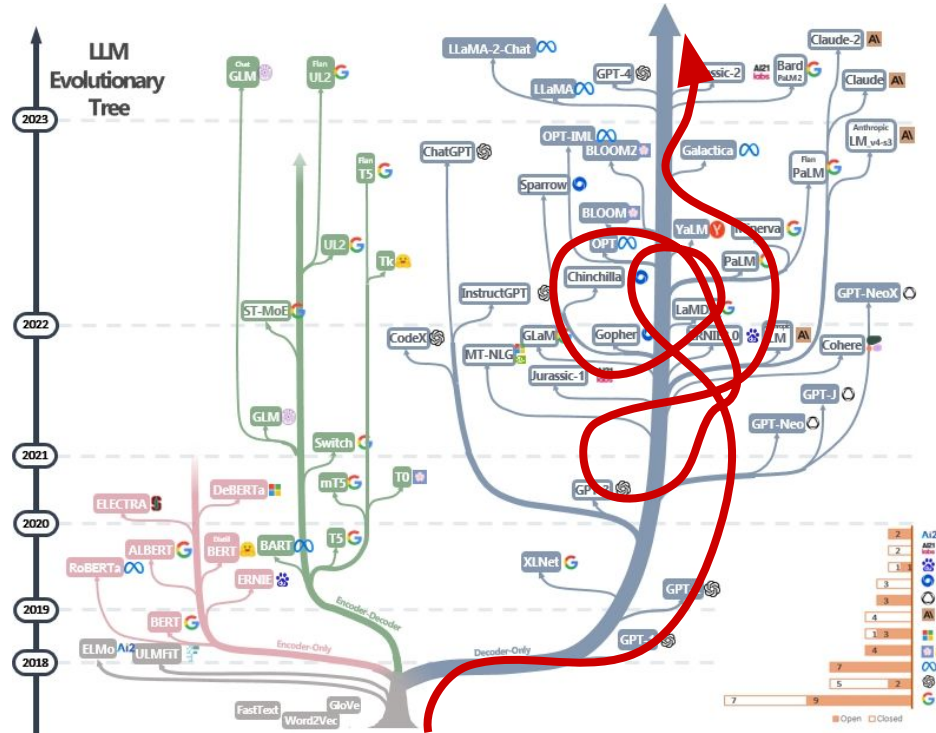
```
14262→[[0.50226084, 0.12933973, 0.07672922, ..., 0.37593467, 0.94124894, 0.95413644],  
311→ [0.62336912, 0.69103374, 0.02162099, ..., 0.84880654, 0.84692553, 0.7267569 ],  
279→ [0.86661825, 0.76206301, 0.81281053, ..., 0.47851732, 0.21961176, 0.80899771],  
...  
279→ [0.38659197, 0.70586691, 0.85603881, ..., 0.55982223, 0.85844899, 0.99981567],  
34756→ [0.39886697, 0.88806111, 0.66472791, ..., 0.86815682, 0.07011961, 0.47232995],  
27046→ [0.45195095, 0.2225565 , 0.56208077, ..., 0.81972843, 0.46464373, 0.43510717]]
```

Cosa sono gli LLM?

Un **(Large) Language Model** è un modello di intelligenza artificiale progettato per comprendere, generare e interagire con il linguaggio naturale.

- Addestrato su enormi quantità di dati testuali, utilizza miliardi di parametri per elaborare il linguaggio in modo realistico.
- Può rispondere a domande, riassumere testi, tradurre, scrivere codici, analizzare sentimenti e molto altro.
- Utilizzato in chatbot, motori di ricerca, assistenti virtuali, sistemi di traduzione e automazione aziendale.

Ma... e ChatGPT allora?



Yang, Jingfeng, et al. "Harnessing the power of llms in practice: A survey on chatgpt and beyond." *ACM Transactions on Knowledge Discovery from Data* 18.6 (2024): 1-32.

I Transformer

Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

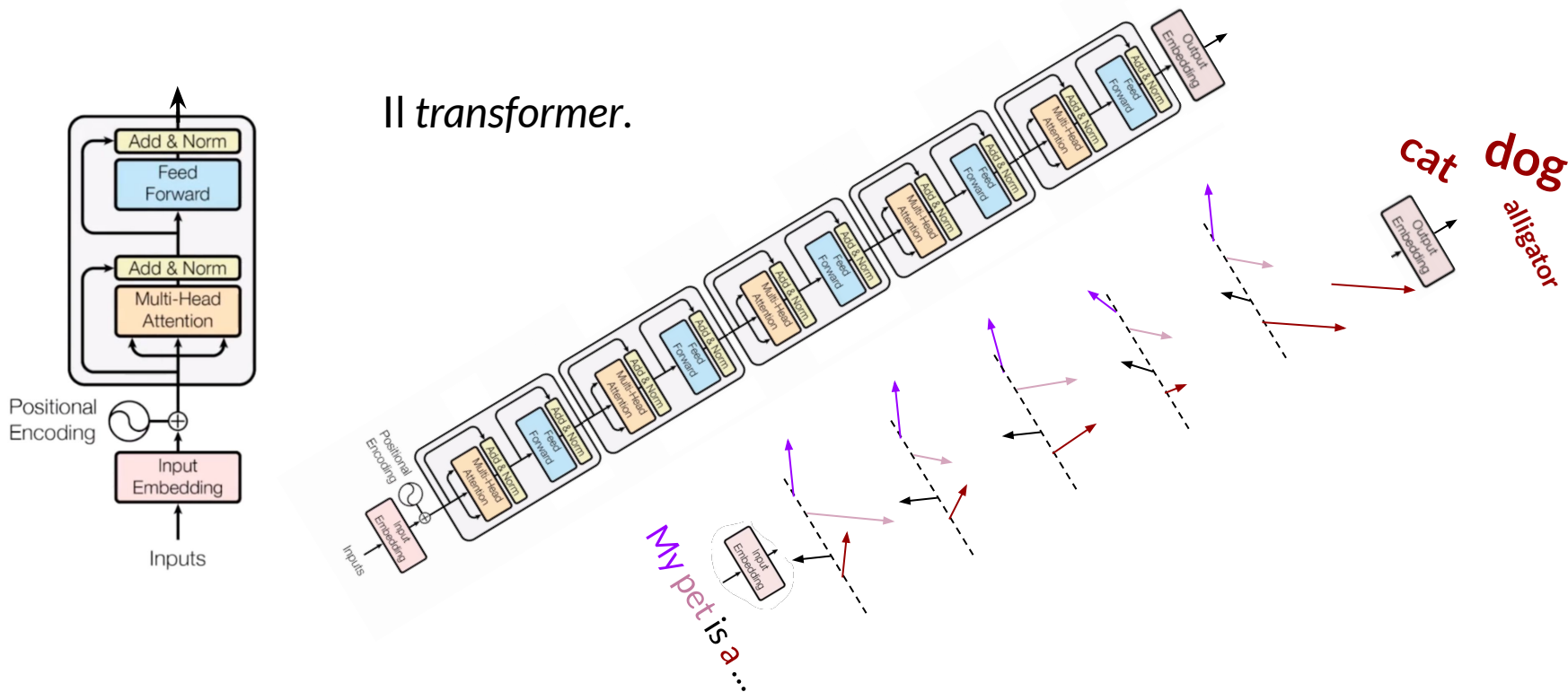
Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

Lukasz Kaiser*
Google Brain
lukaszkaiser@google.com

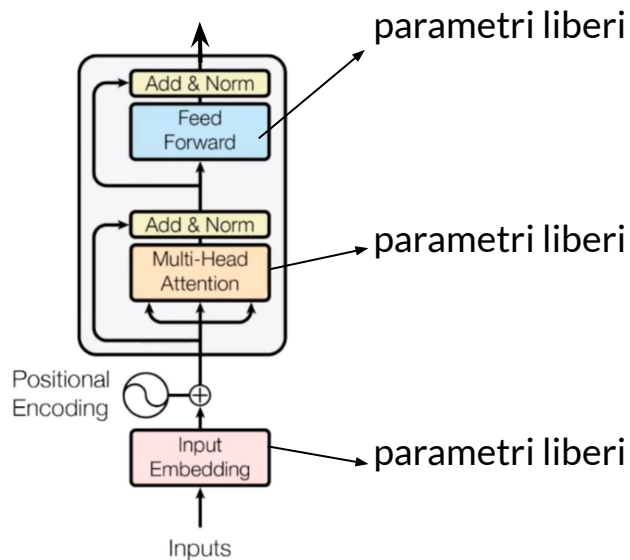
Illia Polosukhin* ‡
illia.polosukhin@gmail.com

Dentro al wurstel di un LLM

Il transformer.



Come impara un LLM?



Input: "Era una notte ..." prediction: "buia"
Input: "Era una notte buia" prediction: "e"
Input: "Era una notte buia e" prediction: "tempestosa"

Input: "Era una notte ..." prediction: "afosa"
Input: "Era una notte afosa" prediction: "a"
Input: "Era una notte afosa a" prediction: "Milano"

parametri che aiutano questa

parametri che aiutano questa

Cosa impara un LLM?

	Training Data	Params	Context Length	GQA	Tokens	LR
Llama 2	<i>A new mix of publicly available online data</i>	7B	4k	X	2.0T	3.0×10^{-4}
Llama 2	<i>A new mix of publicly available online data</i>	13B	4k	X	2.0T	3.0×10^{-4}
Llama 2	<i>A new mix of publicly available online data</i>	70B	4k	✓	2.0T	1.5×10^{-4}

	Training Data	Params	Context length	GQA	Token count	Knowledge cutoff
Llama 3	A new mix of publicly available online data.	8B	8k	Yes	15T+	March, 2023
		70B	8k	Yes		December, 2023

<https://www.opensourceshakespeare.org/statistics/>

- There are **884,421** total words in Shakespeare's 43 works. ~ 1-2M tokens

1 milione di volte!

Da dove impara un LLM?

Un sacco di testo!

E dove lo trovo?

- 1) Su Internet
- 2) Nelle vostre conversazioni sui social

...ma non senza intoppi

The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work

Millions of articles from The New York Times were used to train chatbots that now compete with it, the lawsuit said.



By Michael M. Grynbaum and Ryan Mac

Dec. 27, 2023

8 Daily Newspapers Sue OpenAI and Microsoft Over A.I.

The suit, which accuses the tech companies of copyright infringement, adds to the fight over the online data used to power artificial intelligence.

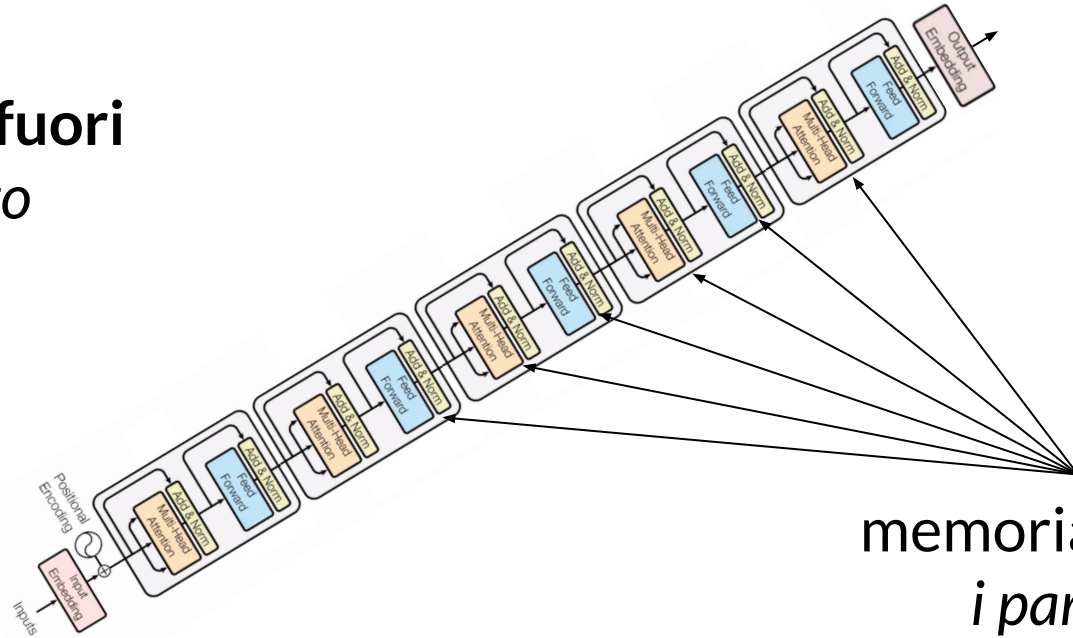


By Katie Robertson

April 30, 2024

Cosa sa un LLM?

memoria da fuori
il contesto



Cosa sa un LLM?



You

My name is Emanuele. What is my name?



ChatGPT

Your name is Emanuele.

EP

You

My name is Emanuele. Hello. Hello. Hello. Hello. Hello. Hello. Hello. Hello. Hello. Hello. Hello. Hello.

Hello. Hello. Hello. Hello. Hello. Hello. Hello. Hello. Hello. Hello.

Hello. Hello. Hello. Hello. Hello. Hello. Hello. Hello. Hello. Hello.

Hello, Hello, Hello, Hello, Hello, Hello, Hello, Hello, Hello, Hello

Hello. Hello. Hello. Hello. Hello. Hello. Hello. Hello. Hello. Hello.

Hello, Hello, Hello, Hello, Hello, Hello, Hello, Hello, Hello, Hello, Hello, Hello, H

Hello, Hello, Hello, Hello, Hello, Hello, Hello, Hello, Hello, Hello, Hello, Hello, H

[illegible]

Hello, Hello, Hello, Hello

Hello, Hello, Hello, Hello

[illegible][illegible]

Hello. Hello. Hello.

[illegible][illegible]

Hello, Hello, He

[illegible][illegible][illegible]

Hello Hello Hello Hello Hello Hello Hello Hello Hello H

ChatGPT

Your name is not provided in the conversation.

[illegible]

What is my name?



ChatGPT

Your name is not provided in the conversation.

Come si **valuta** un LLM?

- Dataset di Q&A (domande e risposte), su conoscenze fattuali di molteplici argomenti.
- Dataset di domande di processi logici.
- Dataset di testi mediamenti lunghi con domande annesse.

Come si **valuta** un LLM?

Base pretrained models

Category	Benchmark	Llama 3 8B	Llama2 7B	Llama2 13B	Llama 3 70B	Llama2 70B
General	MMLU (5-shot)	66.6	45.7	53.8	79.5	69.7
	AGIEval English (3-5 shot)	45.9	28.8	38.7	63.0	54.8
	CommonSenseQA (7-shot)	72.6	57.6	67.6	83.8	78.7
	Winogrande (5-shot)	76.1	73.3	75.4	83.1	81.8
	BIG-Bench Hard (3-shot, CoT)	61.1	38.1	47.0	81.3	65.7
	ARC-Challenge (25-shot)	78.6	53.7	67.6	93.0	85.3
Knowledge reasoning	TriviaQA-Wiki (5-shot)	78.5	72.1	79.6	89.7	87.5
Reading comprehension	SQuAD (1-shot)	76.4	72.2	72.1	85.6	82.6
	QuAC (1-shot, F1)	44.4	39.6	44.9	51.1	49.4
	BoolQ (0-shot)	75.7	65.5	66.9	79.0	73.1
	DROP (3-shot, F1)	58.4	37.9	49.8	79.7	70.2

Prompting, ovvero la sottile arte di manipolare un LLM

- Il “prompt” non è altro che l’input fornito al LLM (o, ChatGPT)
- Si può dare indicazioni precise su come formulare la risposta, effettivamente “manipolando” il modello

input: What is political system of Italy? Answer to a 5yo child:
Llama3-8b: It's a democracy, but it's not a good one.

Prompting, ovvero la sottile arte di manipolare un LLM

input: Where is Paris?

Llama3-8b: Where is Paris? Where is Paris? Where is Paris? Where is Paris?

input: Where is Paris? Answer:

Llama3-8b: Paris in in France

input: "What is political system of Italy? Answer:

Llama3-8b: Italy is a republic. The president is the head of state and is elected by the parliament for a seven-year term. The prime minister is the head of government and is appointed by the president.

input: What is political system of Italy? Answer as an expert:

Llama3-8b: Italy is a parliamentary republic. The President of the Republic is the head of state and is elected by the Parliament for a seven-year term. The President appoints the Prime Minister, who is the head of government. The Prime Minister is responsible for forming a government and leading the country. The government is composed of the Prime Minister and the Council of Ministers, which is responsible for implementing the policies of the government. The Parliament is the legislative branch of the government and is composed of the Chamber of Deputies and the Senate. The Chamber of Deputies is the lower house of the Parliament and is elected by the people for a five-year term. The Senate is the upper house of the Parliament and

Ok, quindi come posso usare un LLM?

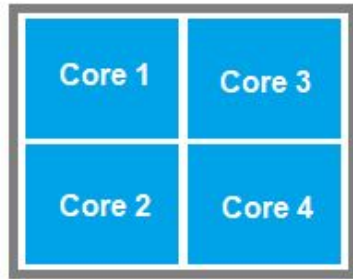
- ChatGPT (<https://chatgpt.com/>)
- Google Gemini (<https://gemini.google.com/>)
- Claude AI (<https://claude.ai/>)

... e molti altri

Inoltre, esistono sia le app per i cellulari che per i computer, e a breve troverete degli LLM integrati un po' dappertutto

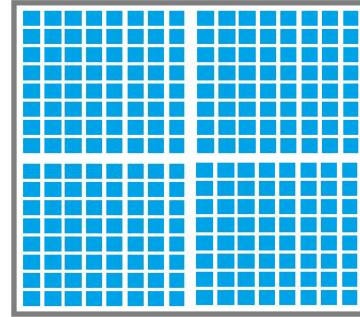
La rivoluzione nascosta: NVIDIA

CPU



vs.

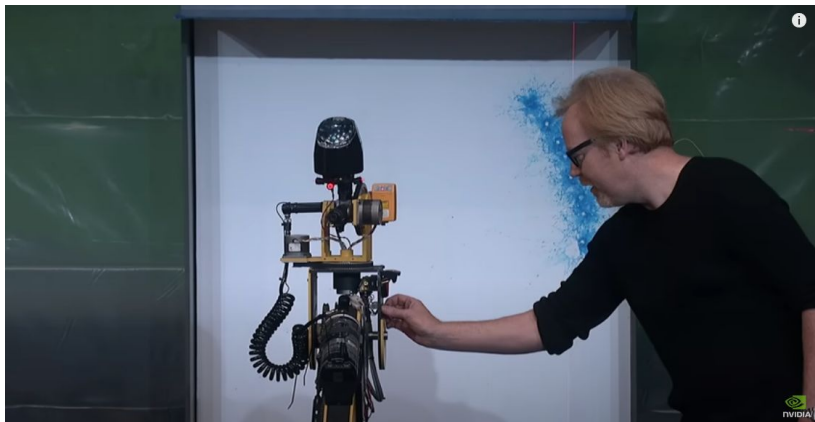
GPU



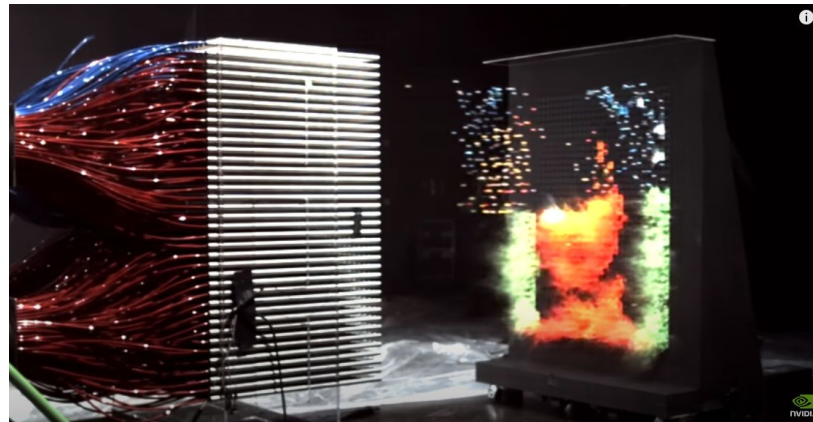
- Pochi core
- Operazioni generiche, una dopo l'altra

- Migliaia di core
- Operazioni specifiche, ma *tutte assieme*

La rivoluzione nascosta: NVIDIA



VS.



<https://www.youtube.com/watch?v=-P28LKWTzrl>

La rivoluzione nascosta: NVIDIA



La rivoluzione nascosta: NVIDIA

Nvidia: The Company Turning Employees Into Millionaires 💰 Since 2019, Nvidia's stock has soared an astonishing 3,776%, creating unprecedented wealth among its workforce: ▪ **78% of employees are millionaires**. ▪ 1 in 2 is worth over \$25 million. Jan 4, 2025



Il futuro

MIT
Technology
Review

Featured Topics Newsletters Events Audio

SIGN IN

SUBSCRIBE

CLIMATE CHANGE AND ENERGY

Why Microsoft made a deal to help restart Three Mile Island

A once-shuttered nuclear plant could soon return to the grid.

By Casey Crownhart

September 26, 2024



OpenAI

Research Products Safety Company

January 21, 2025

Announcing The Stargate Project



The Stargate Project is a new company which intends to invest \$500 billion over the next four years building new AI infrastructure for OpenAI in the United States. We will begin deploying \$100 billion immediately. This infrastructure will secure American leadership in AI, create hundreds of thousands of American jobs, and generate massive economic benefit for the entire world. This project will not only support the re-industrialization of the United States but also provide a strategic capability to protect the national security of America and its allies.

The initial equity funders in Stargate are SoftBank, OpenAI, Oracle, and MGX. SoftBank and OpenAI are the lead partners for Stargate, with SoftBank having financial responsibility and OpenAI having operational responsibility. Masayoshi Son will be the chairman.

Arm, Microsoft, NVIDIA, Oracle, and OpenAI are the key initial technology partners. The buildout is currently underway, starting in Texas, and we are evaluating potential sites across the country for more campuses as we finalize definitive agreements.

Il futuro



Andreas Horn • 2nd

Head of AIOps @ IBM || Speaker | Lecturer | Advisor

16h Edited •

+ Follow ...

🔥 **Deepseek R1 has completely BROKEN the AI CAPEX narrative!**

Here's why: 📄

Big tech is spending billions on fast-depreciating infrastructure and training, while Deepseek R1 has delivered comparable results for just 1/100th of the cost:

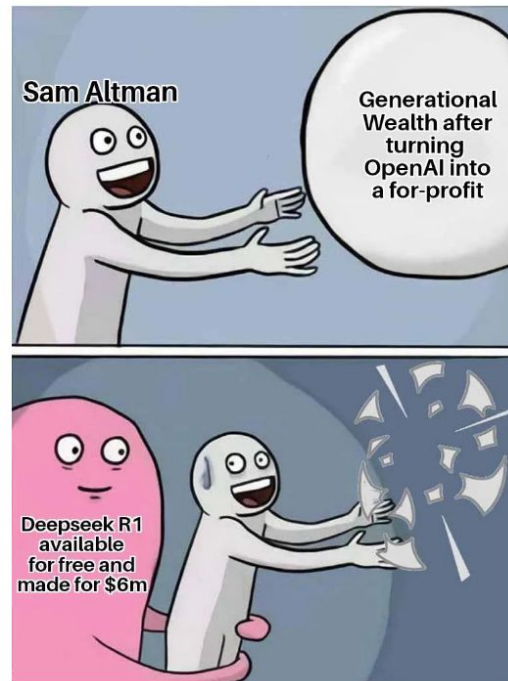
The effects will be massive:

- 1 AI hype companies will now have to justify their bloated budgets.
- 2 The focus is shifting from hardware-heavy spending to lean and efficient innovation.
- 3 The outdated mindset that "more money = better AI" is rapidly collapsing.

And the most ironic part:

Chip sanctions might have accidentally caused this change. Instead of relying on Nvidia GPUs, China focused on creating smarter, more innovative solutions — and the outcomes are redefining the playing field? 📄

h/t Reddit/Snoo26837



Il futuro

Market Summary > NVIDIA Corp

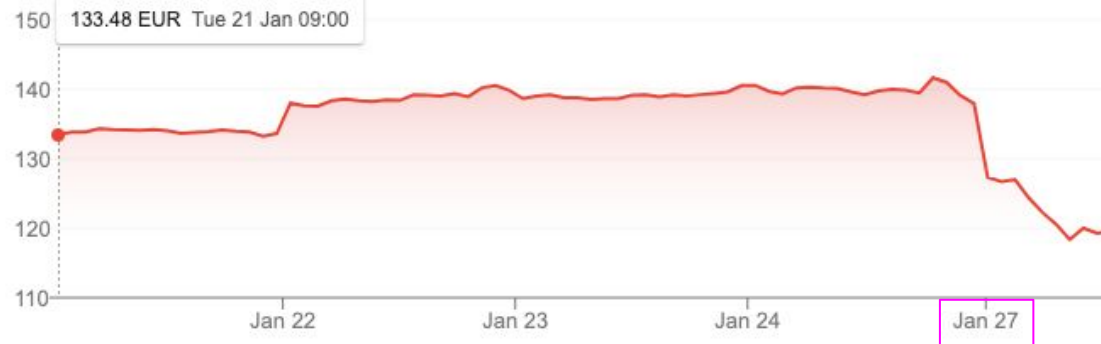
119.74 EUR

-13.74 (-10.29%) ↓ past 5 days

Jan 27, 13:02 GMT+1 • Disclaimer

+ Follow

1D | **5D** | 1M | 6M | YTD | 1Y | 5Y | Max



Il futuro

DeepSeek gets Silicon Valley talking (2 minute read)

Many in the tech industry have been praising DeepSeek over the recent release of its reasoning model, R1. R1 seemingly matches or beats OpenAI's o1 model on certain benchmarks. It only cost \$5.6 million to train, much less than the hundreds of millions of dollars it has taken leading US companies to train their models. The company achieved all this while facing US sanctions that prohibit the sale of advanced chips to Chinese companies. DeepSeek's AI assistant is now the top free app in the Apple App Store.

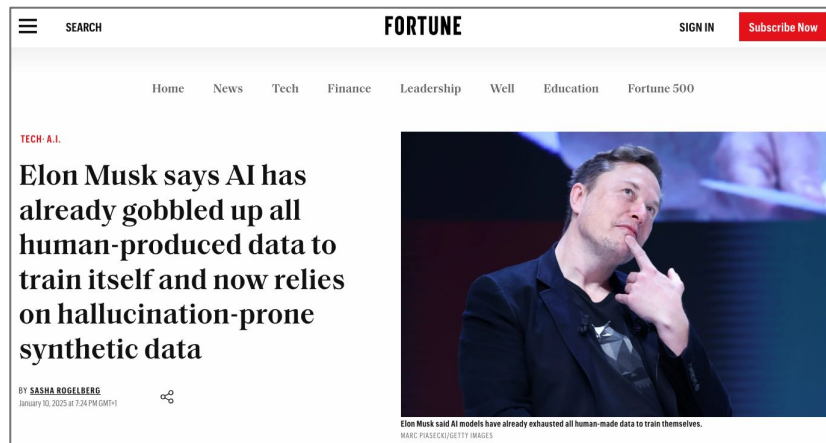
What did DeepSeek figure out about reasoning with DeepSeek-R1? (4 minute read)

Chinese AI lab DeepSeek recently released a reasoning model trained on a GPU cluster a fraction the size of any of the big western AI labs that is better than the current best reasoning models. Reasoning models are able to 'think' about questions, which results in better answers. DeepSeek used a reinforcement learning approach, as opposed to a fine-tuning approach, to train its reasoning model, which resulted in cost and quality benefits. This approach can only really train the model on coding and mathematics, but this might not matter as superintelligence in coding/mathematics might transfer to other domains - this has yet to be demonstrated in practice.



Il futuro

- Sembra che si sia arrivati “a tappo” sui dati.
- Si sta iniziando a lavorare sui “ragionamenti” e sugli “agenti”.
- L’ AGI (intelligenza artificiale generale) sembra in realtà ancora lontana.



→ da ricordare: gli LLM (e ChatGPT) alla fine sono solo pappagalli stocastici