NORDCOM

Innovare nell’era dell’artificial intelligence

Francesco M. Sacco
Università dell’Insubria e SDA Bocconi, StartUp Economy, Kebula, S3K

TECH

Elon Musk says global race for A.I. will be the most likely cause of World War III

PUBLISHED MON, SEP 4 2017•5:52 AM EDT | UPDATED MON, SEP 4 2017•6:46 AM EDT

Ryan Browne
@RYAN_BROWNE_

KEY POINTS

- Elon Musk said Monday that artificial intelligence could be the "most likely" cause of a third world war
- Russian President Vladimir Putin said Friday that the first global leader in AI would "become the ruler of the world"

DAVOS WEF

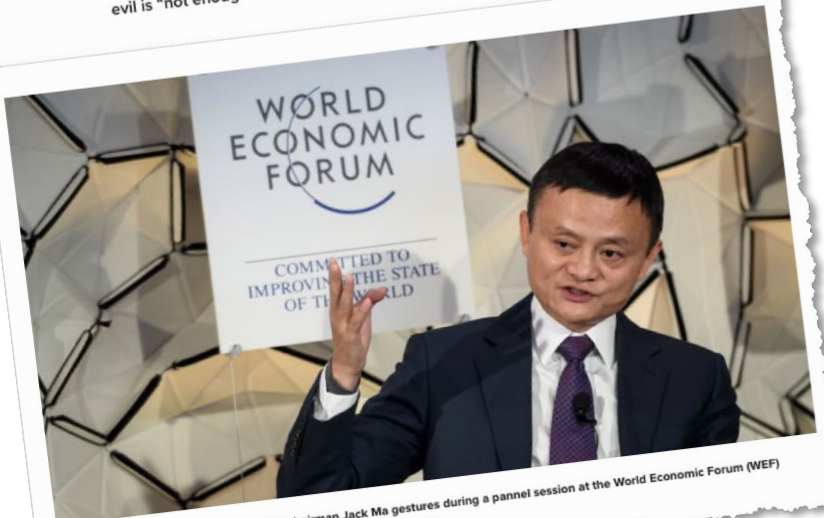
Alibaba's Jack Ma suggests technology could result in a new world war

PUBLISHED WED, JAN 23 2019•9:29 AM EST | UPDATED FRI, JAN 25 2019•9:31 AM EST

Ryan Browne
@RYAN_BROWNE_

KEY POINTS

- Ma said that history shows technological innovation is a primary factor behind global conflicts.
- It's not the first time the business leader has made such a warning — he made a similar comment in 2017.
- In reference to the Google motto "don't be evil," Ma said that tech firms not being evil is "not enough."



...man Jack Ma gestures during a pannel session at the World Economic Forum (WEF)

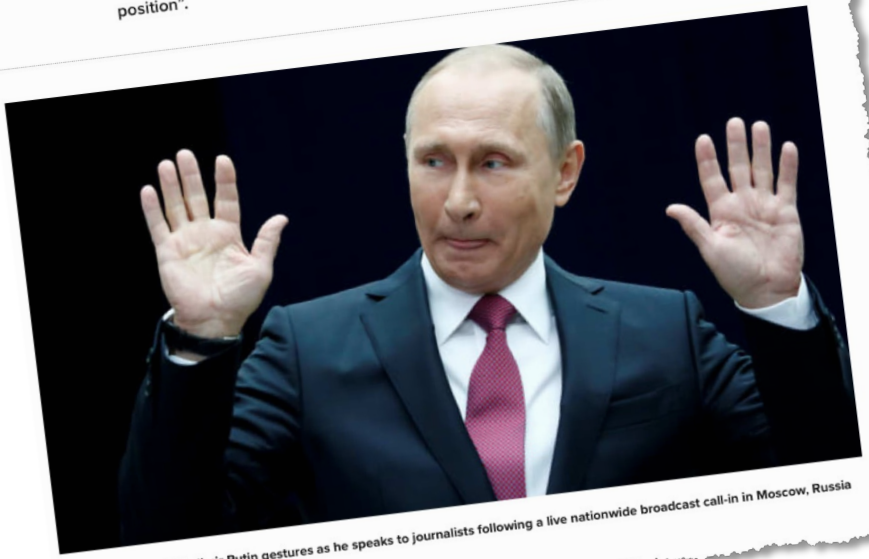
Putin: Leader in artificial intelligence will rule world

PUBLISHED MON, SEP 4 2017-2:33 AM EDT | UPDATED MON, SEP 4 2017-2:33 AM EDT

AP

KEY POINTS

- Putin says that whoever reaches a breakthrough in developing artificial intelligence will come to dominate the world.
- Putin warned that "it would be strongly undesirable if someone wins a monopolist position".



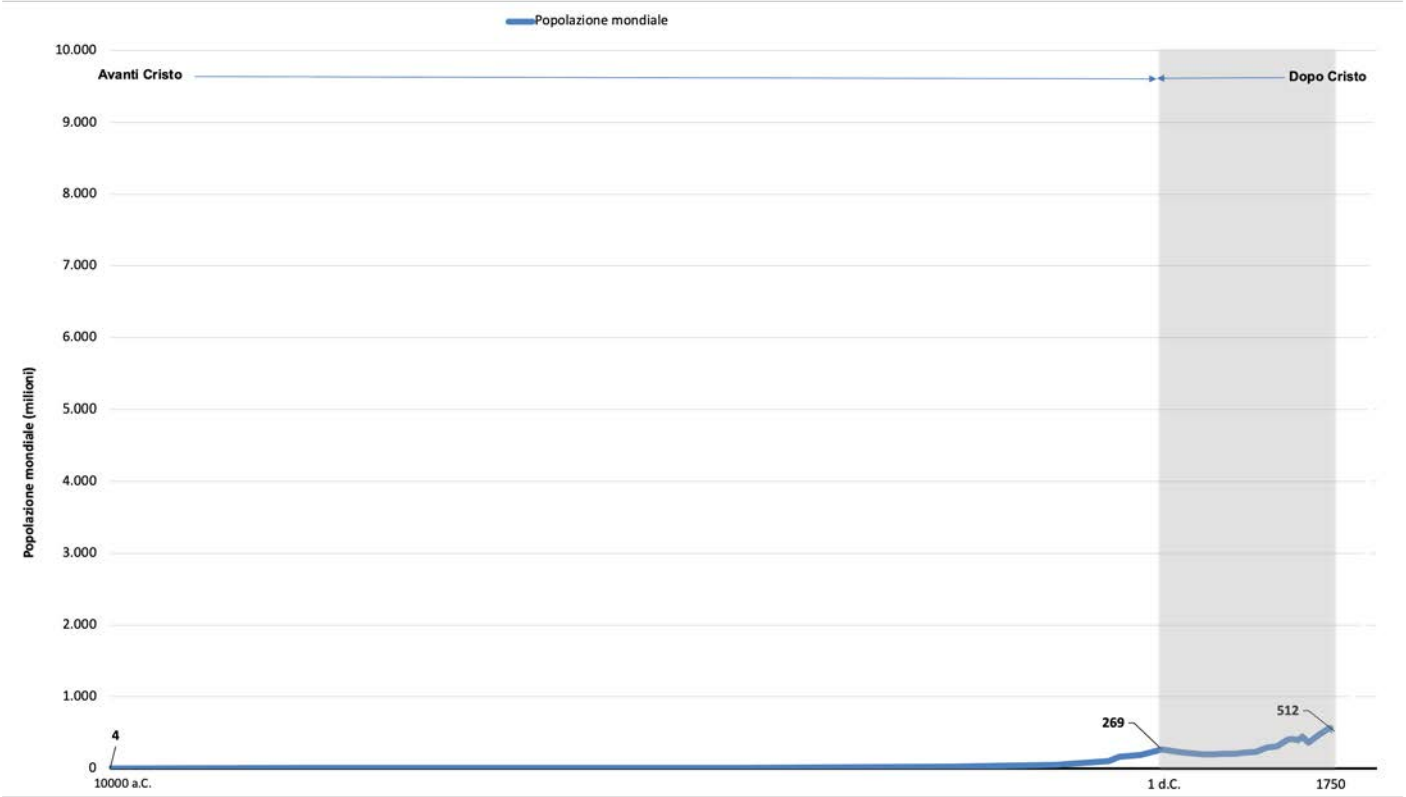
Russian President Vladimir Putin gestures as he speaks to journalists following a live nationwide broadcast call-in in Moscow, Russia June 15, 2017.

Kapyukhin | Reuters

COME É POSSIBILE?

SIAMO IN UN MONDO IN CRESCITA ESPONENZIALE

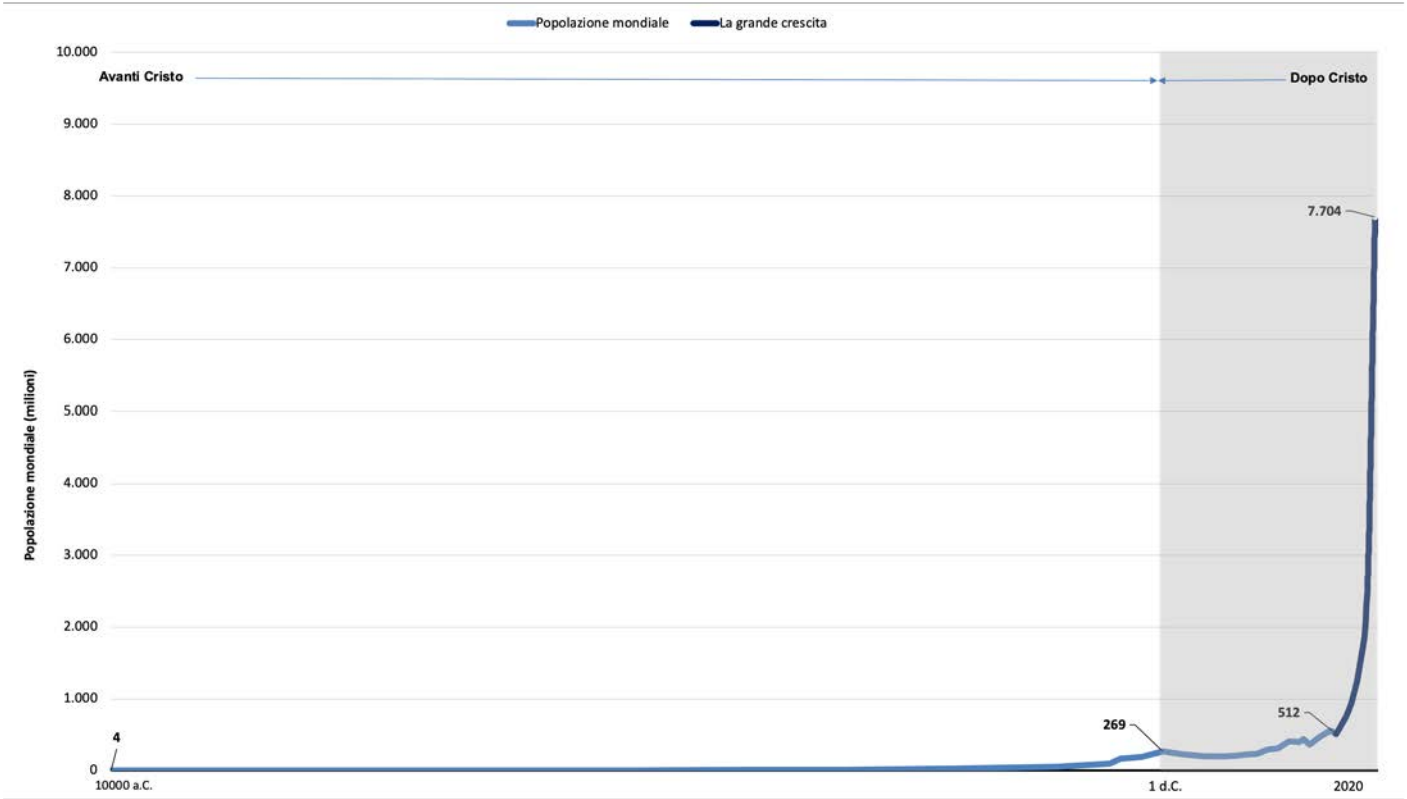
Per popolazione…



Fonti: Ns. elaborazioni su media delle ricostruzioni storiche fino al 1950; dal 1951 dati e previsioni ONU

SIAMO IN UN MONDO IN CRESCITA ESPONENZIALE

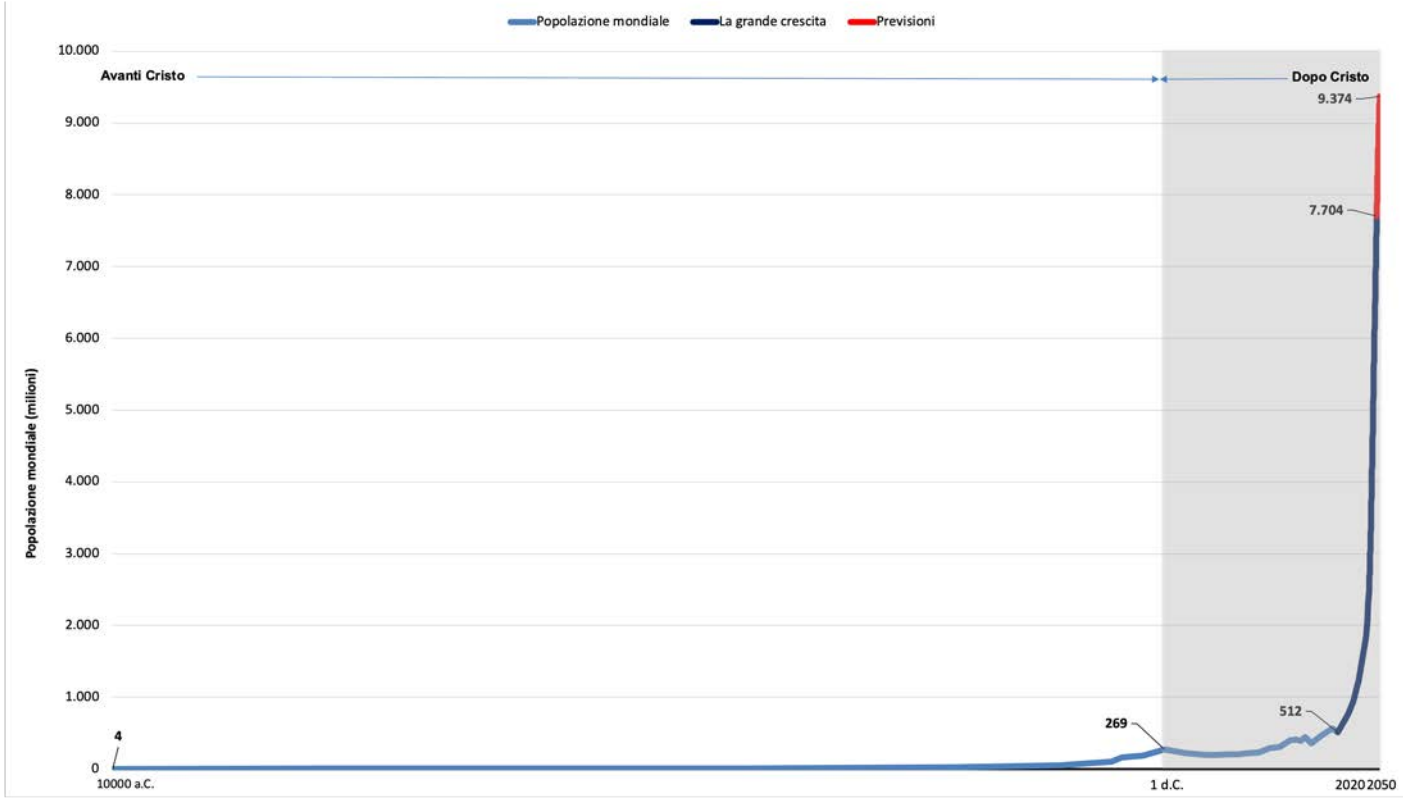
Per popolazione…



Fonti: Ns. elaborazioni su media delle ricostruzioni storiche fino al 1950; dal 1951 dati e previsioni ONU

SIAMO IN UN MONDO IN CRESCITA ESPONENZIALE

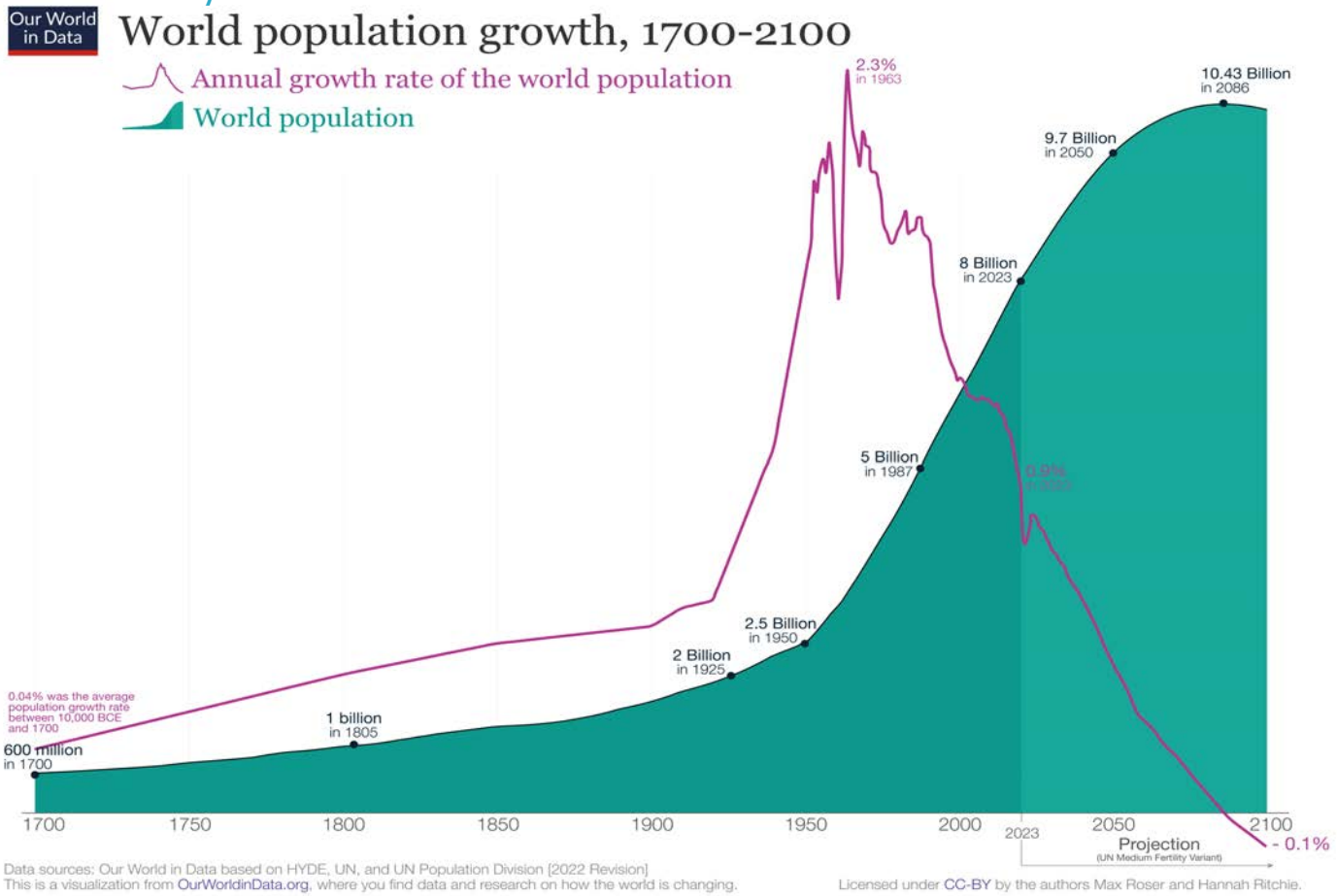
Per popolazione…



Fonti: Ns. elaborazioni su media delle ricostruzioni storiche fino al 1950; dal 1951 dati e previsioni ONU

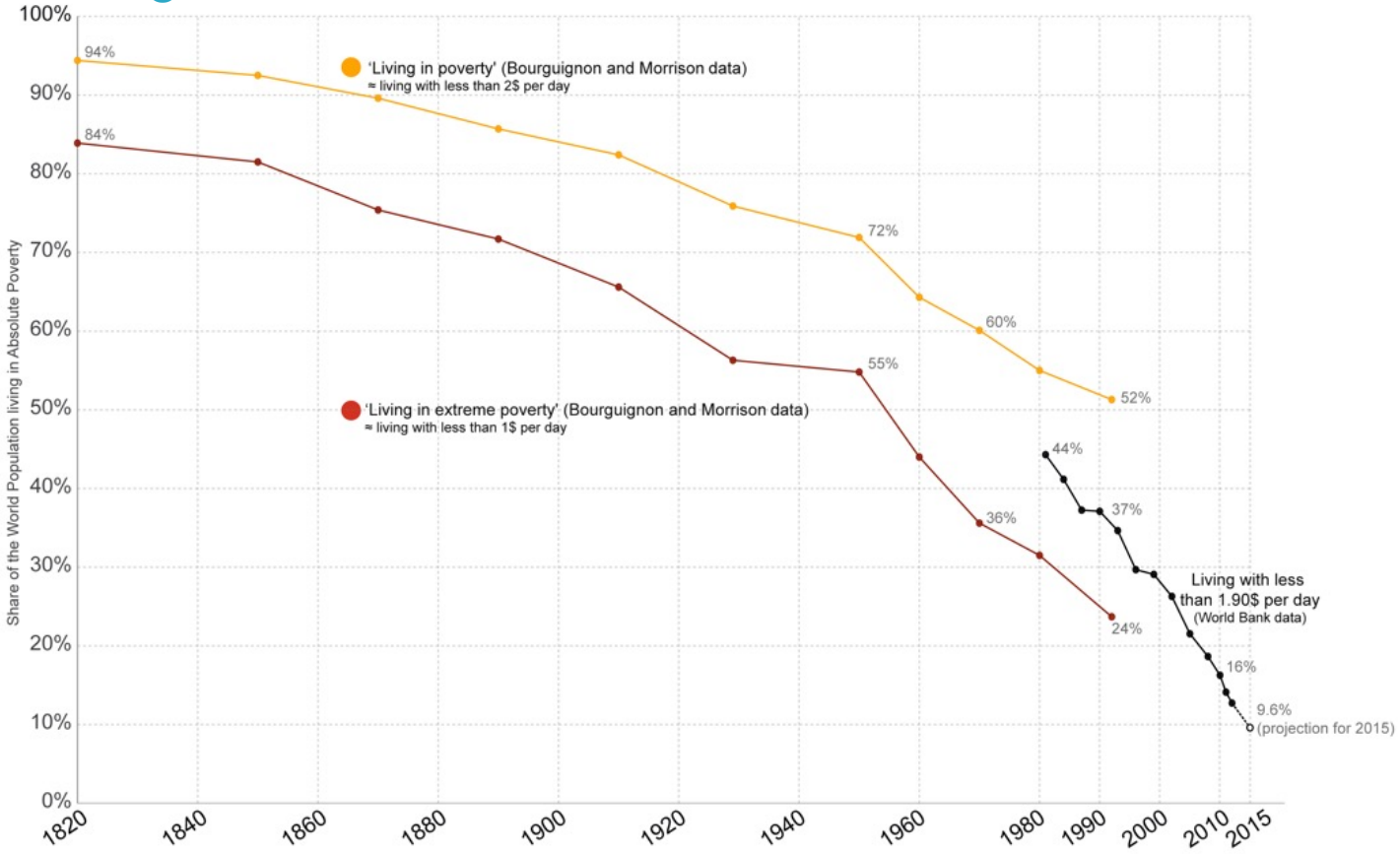
SIAMO IN UN MONDO IN CRESCITA ESPONENZIALE

Per popolazione (revisione 2022)…



SIAMO IN UN MONDO IN CRESCITA ESPONENZIALE

Ma anche in un mondo migliore in termini relativi…

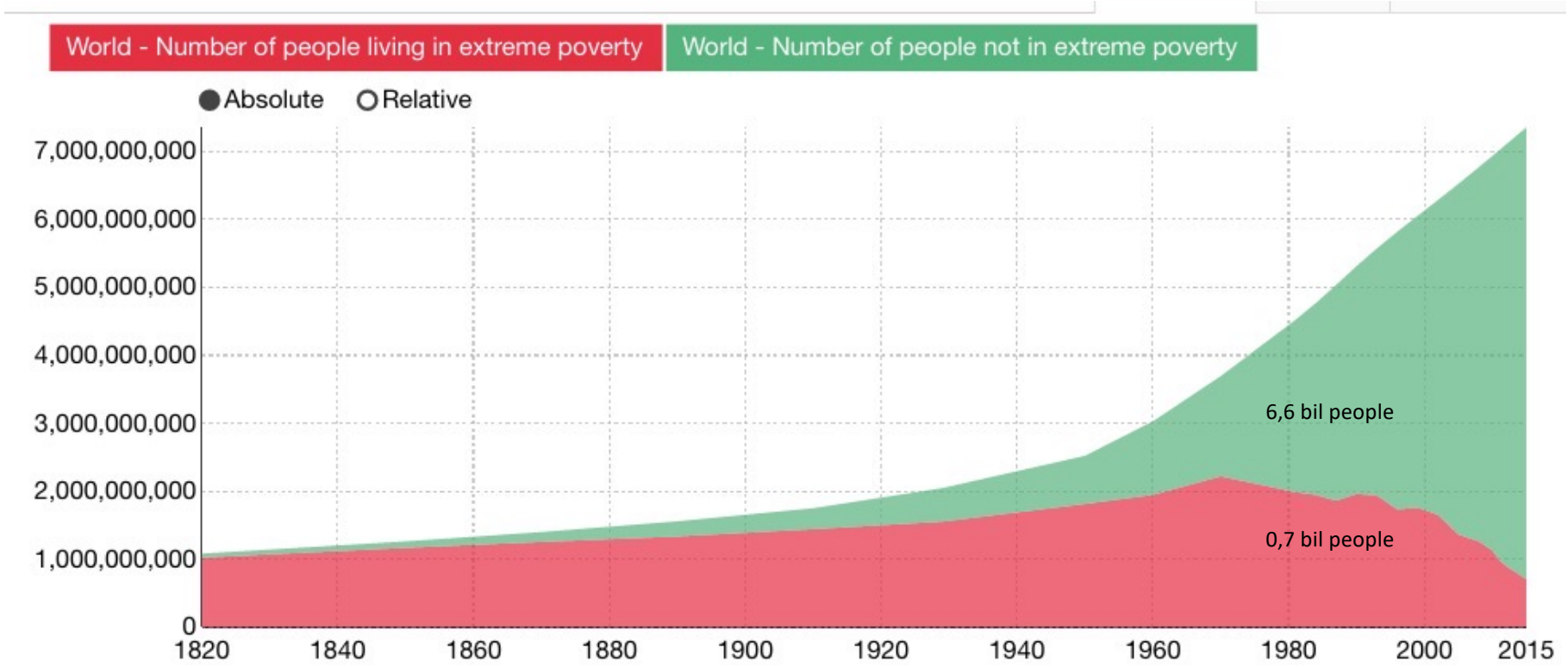


Data sources: 1820-1992 Bourguignon and Morrison (2002) - Inequality among World Citizens, In The American Economic Review; 1981-2015 World Bank (PovcalNet)

The interactive data visualisation is available at OurWorldinData.org. There you find the raw data and more visualisations on this topic.

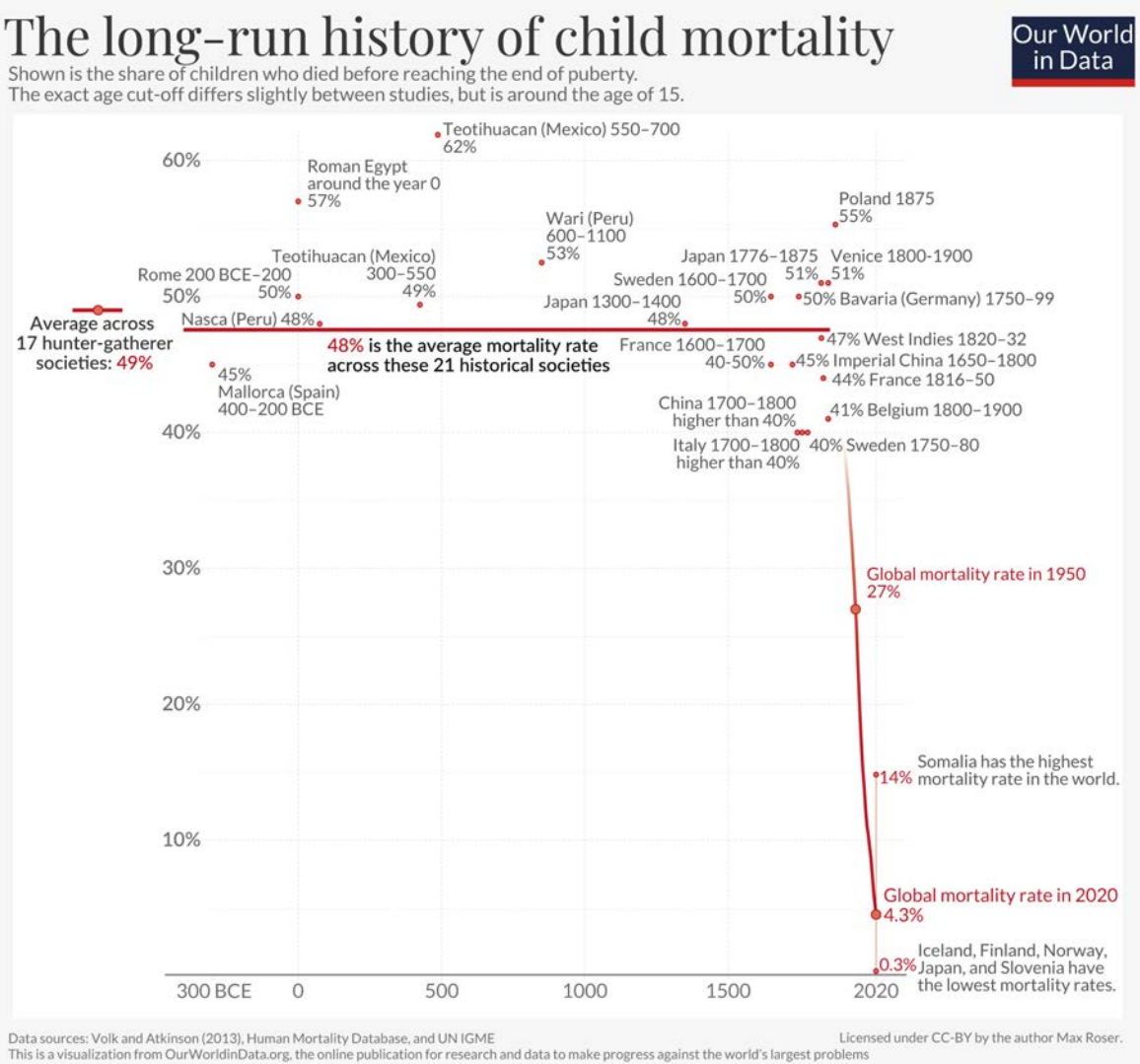
SIAMO IN UN MONDO IN CRESCITA ESPONENZIALE

…e in termini assoluti



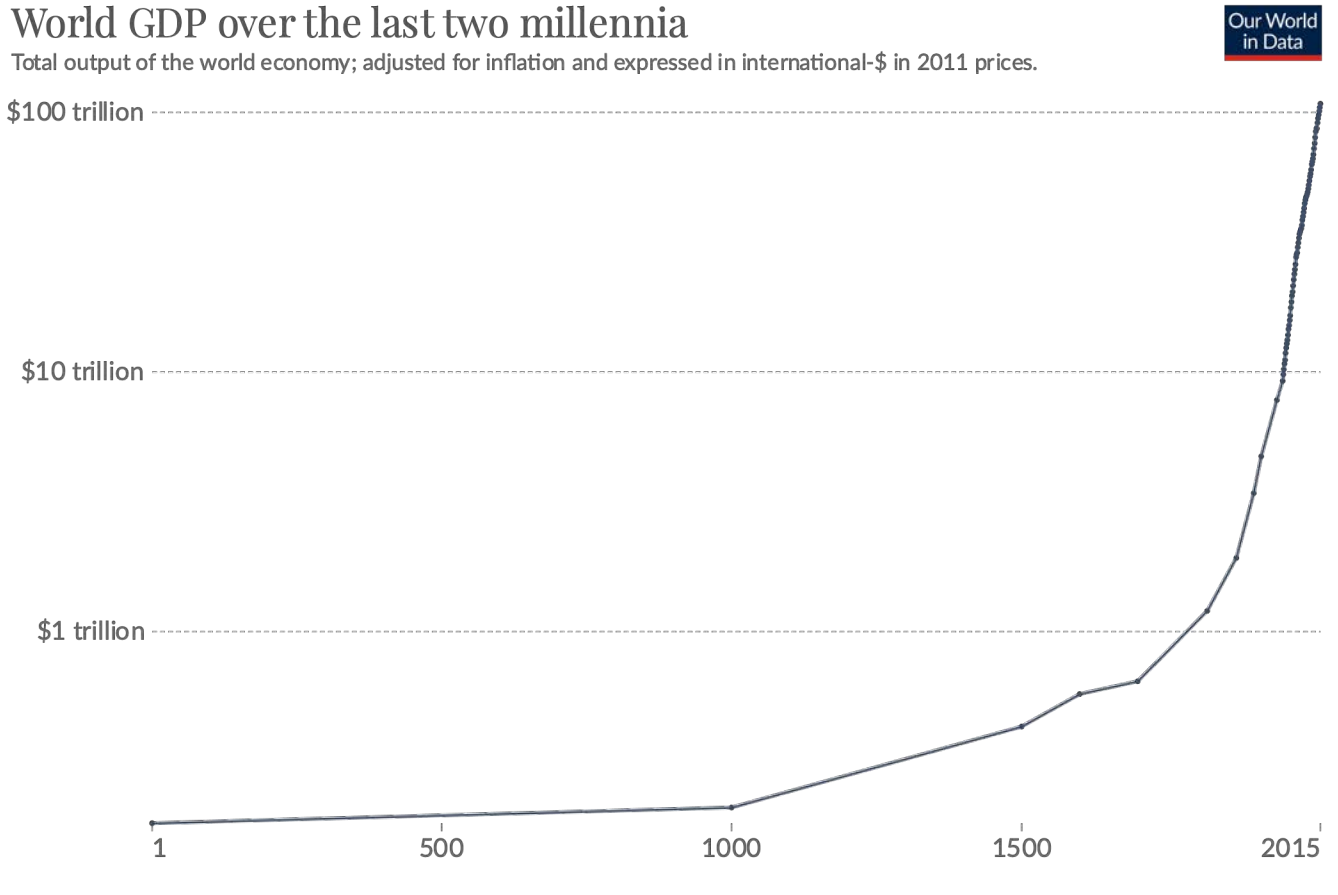
SIAMO IN UN MONDO IN CRESCITA ESPONENZIALE

Anche grazie ai progressi della medicina…



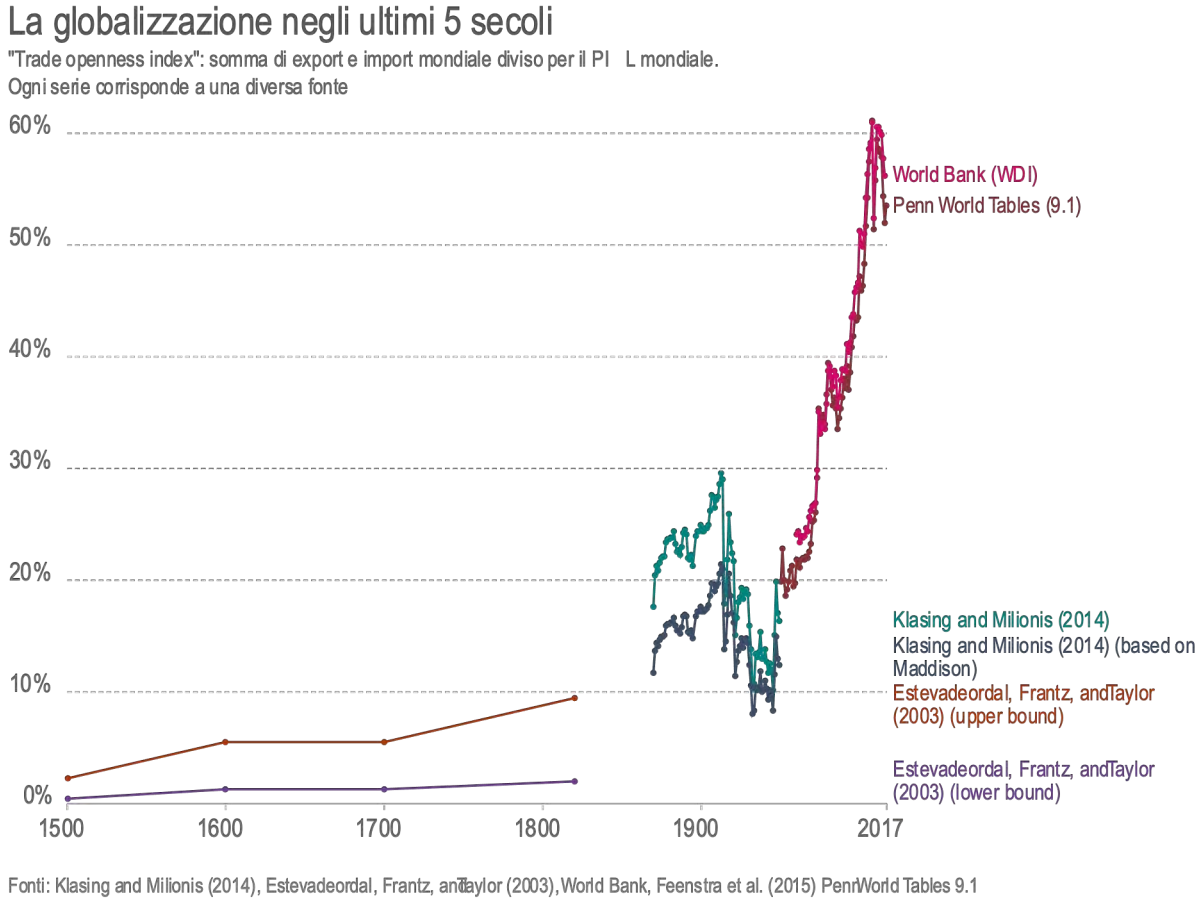
UN MONDO IN CRESCITA ESPONENZIALE

Per PIL…



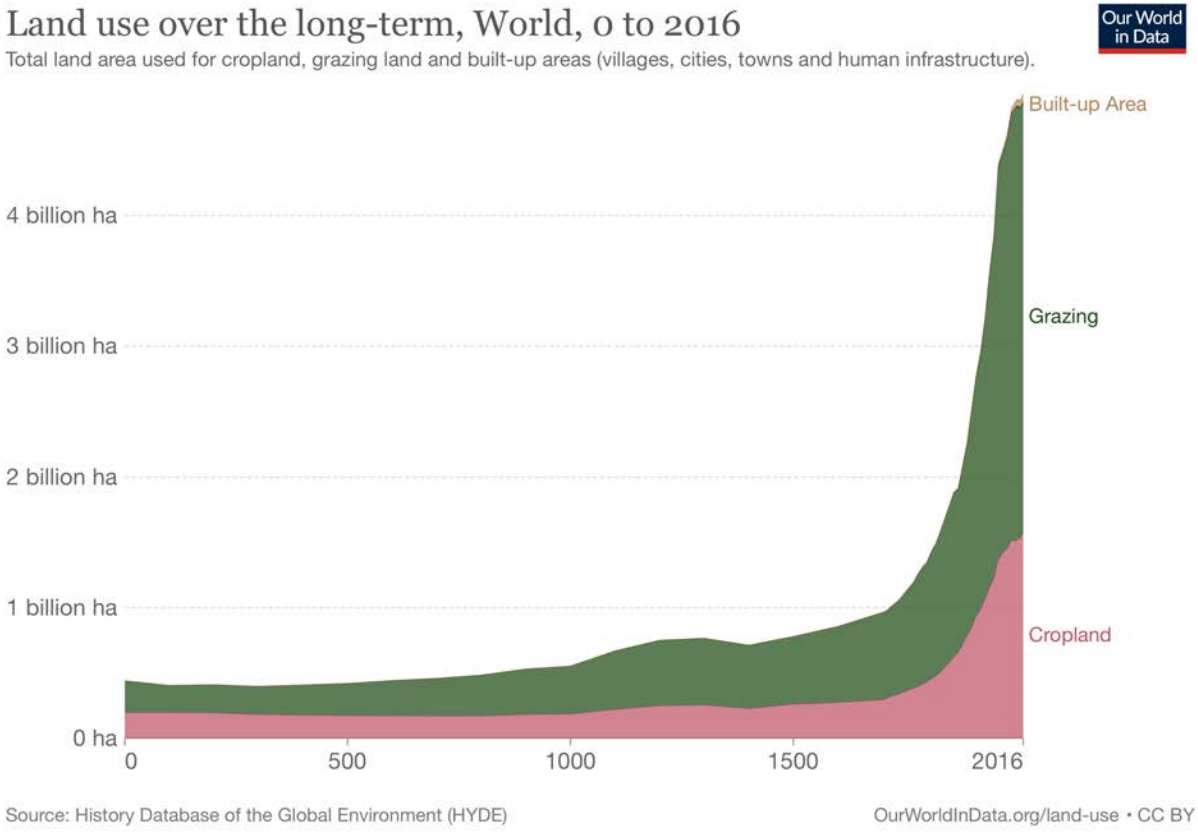
SIAMO IN UN MONDO IN CRESCITA ESPONENZIALE

Per globalizzazione…



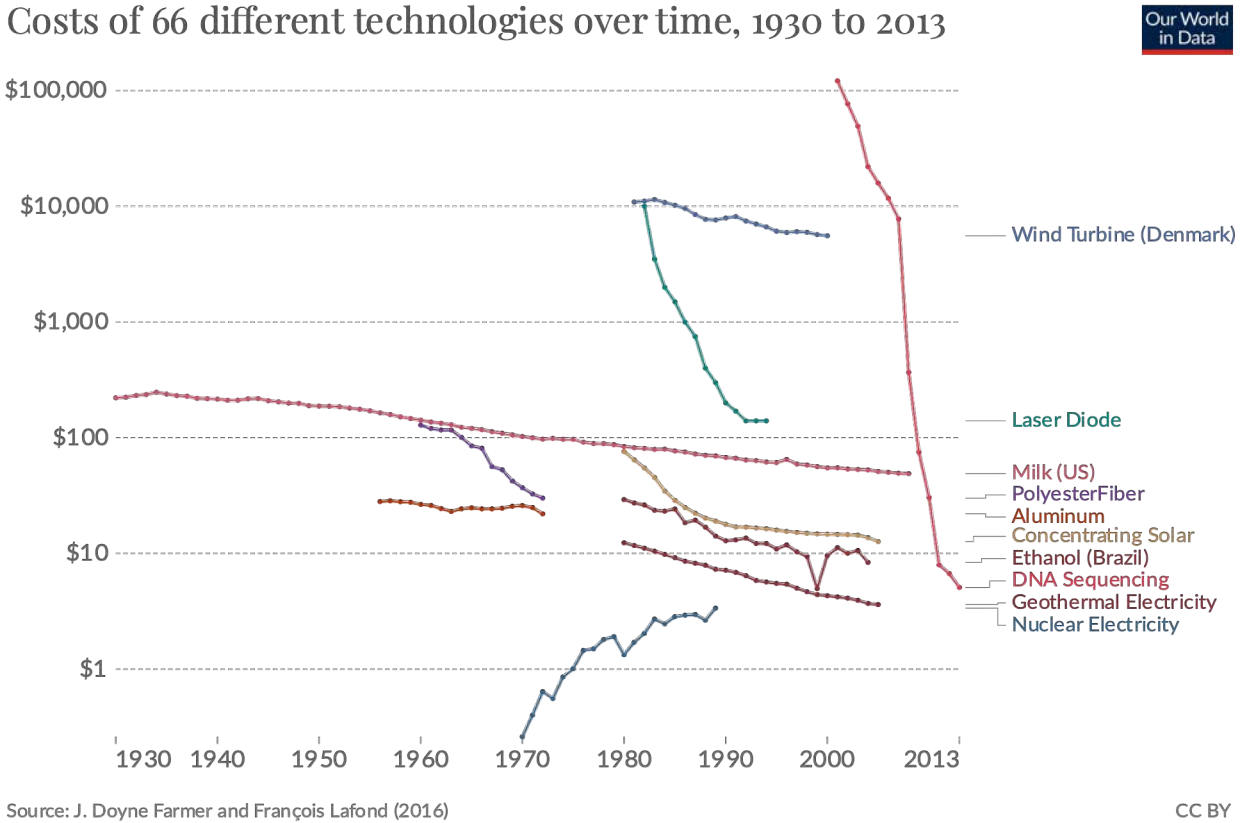
SIAMO IN UN MONDO IN CRESCITA ESPONENZIALE

Per uso del terreno…



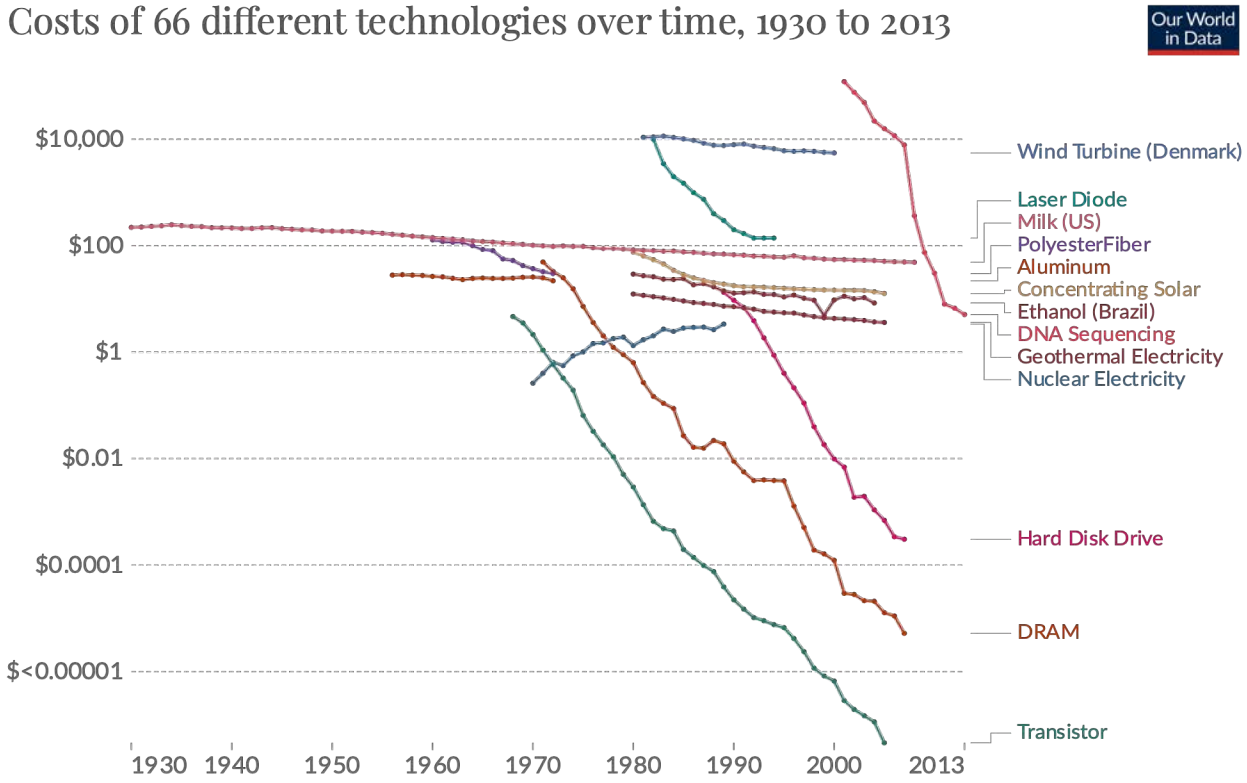
SIAMO IN UN MONDO IN CRESCITA ESPONENZIALE

Per calo dei costi della tecnologia …



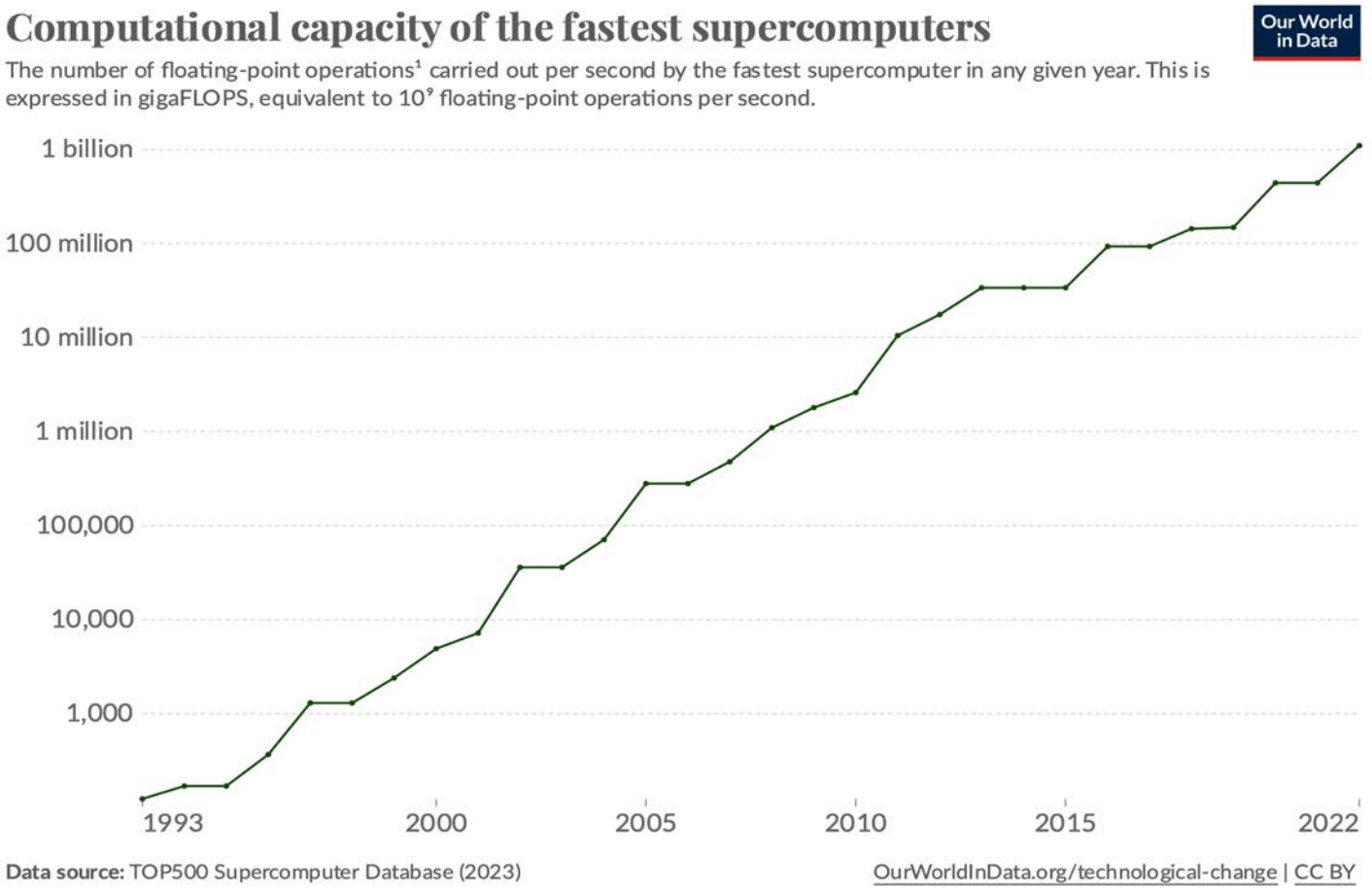
SIAMO IN UN MONDO IN CRESCITA ESPONENZIALE

Per calo dei costi della tecnologia …

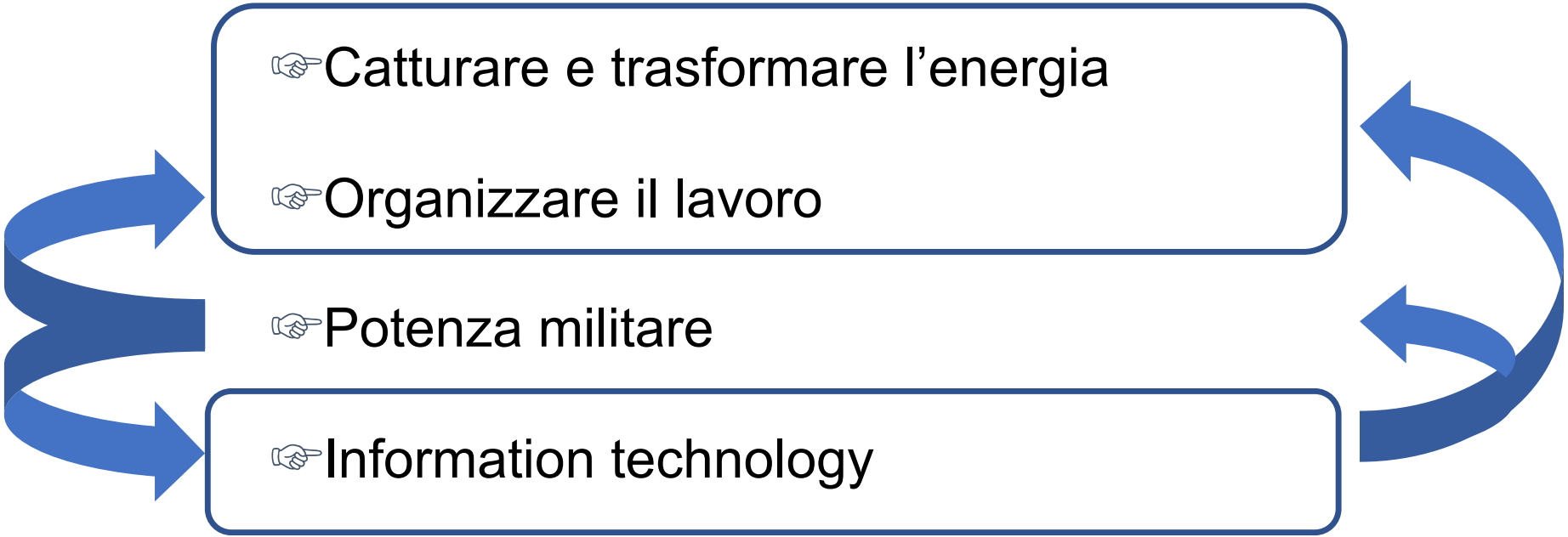


SIAMO IN UN MONDO IN CRESCITA ESPONENZIALE

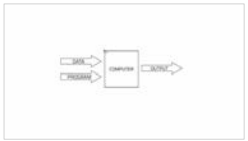
Per potenza computazionale…

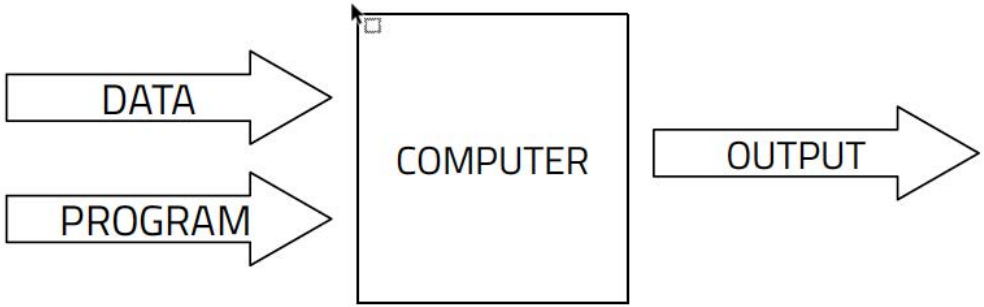


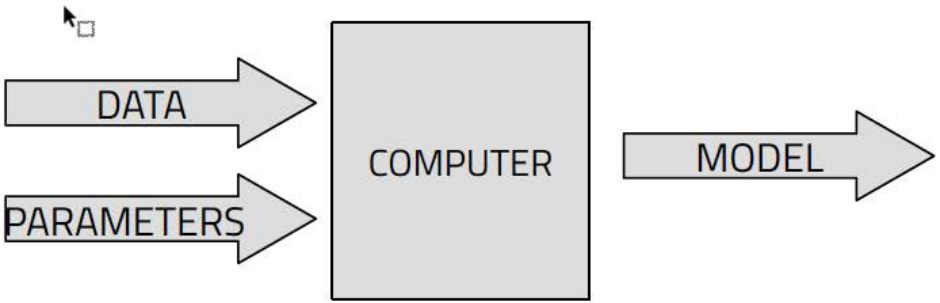
Come è stato possibile?

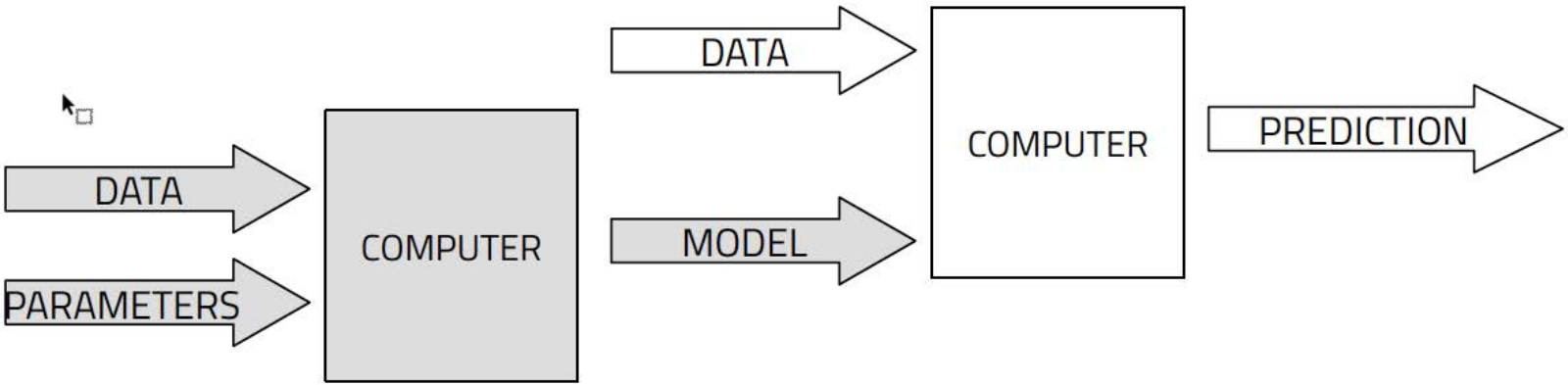

☞Catturare e trasformare l’energia
☞Organizzare il lavoro
☞Potenza militare
☞Information technology

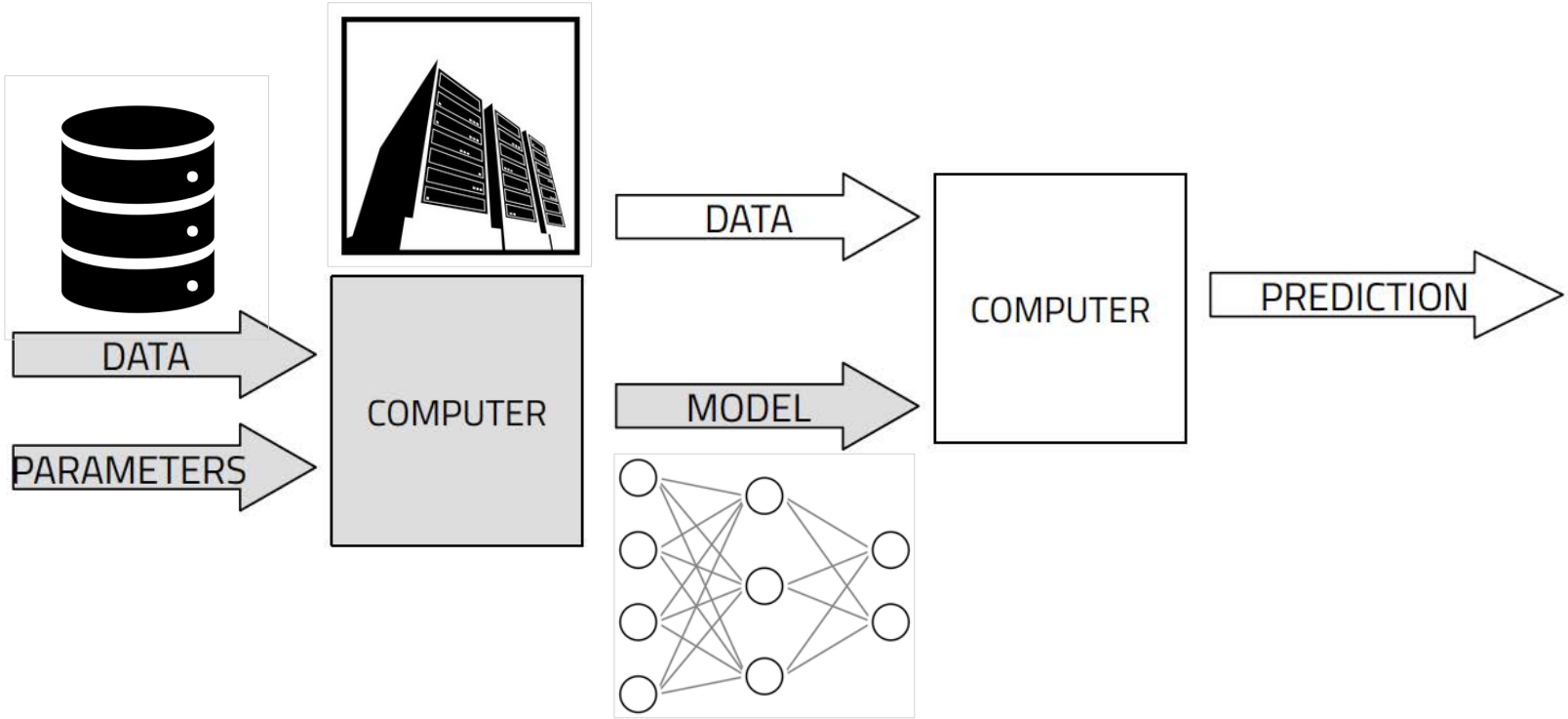
- ☞Catturare e trasformare l’energia
- ☞Organizzare il lavoro
- ☞Potenza militare
- ☞Information technology




DATA
PROGRAM
COMPUTER
OUTPUT


DATA
PARAMETERS
COMPUTER
MODEL


DATA
PARAMETERS
COMPUTER
DATA
MODEL
COMPUTER
PREDICTION


DATA
PARAMETERS
COMPUTER
DATA
MODEL
COMPUTER
PREDICTION

L'addestramento di GPT-3 è stato 45 TB di testo

Tanti dati…

GPT-3 Training data

Dataset	Content	Raw size (tokens, ~words, in billions)	Weight (in training mix)	Composition (actual % of total content mass)	Amplification (or suppression)
Common Crawl	web corpus, incl. images, links = 3.2bn pages	410	60%	82%	-0.27
WebText2	external reddit links = 45m high quality pages	19	20%	4%	+5.00
Books1 & Books 2	selection of 200k books	67	15%	13%	+0.15
Wikipedia	knowledge base in English lang.	3	5%	1%	+5.00

Thompson's analysis of the individual sources of GPT-3 training data

CommonCrawl	
Google Patents	0.48%
The New York Times	0.06%
Los Angeles Times	0.06%
The Guardian	0.06%
Public Library of Science	0.06%
Forbes	0.05%
Huffington Post	0.05%
Patents.com	0.05%
Scribd	0.04%
Other	99.09%

WebText2 (Reddit links)	
Google	3.4%
Archive	1.3%
Blogspot	1.0%
GitHub	0.9%
The New York Times	0.7%
Wordpress	0.7%
Washington Post	0.7%
Wikia	0.7%
BBC	0.7%
Other	89.9%

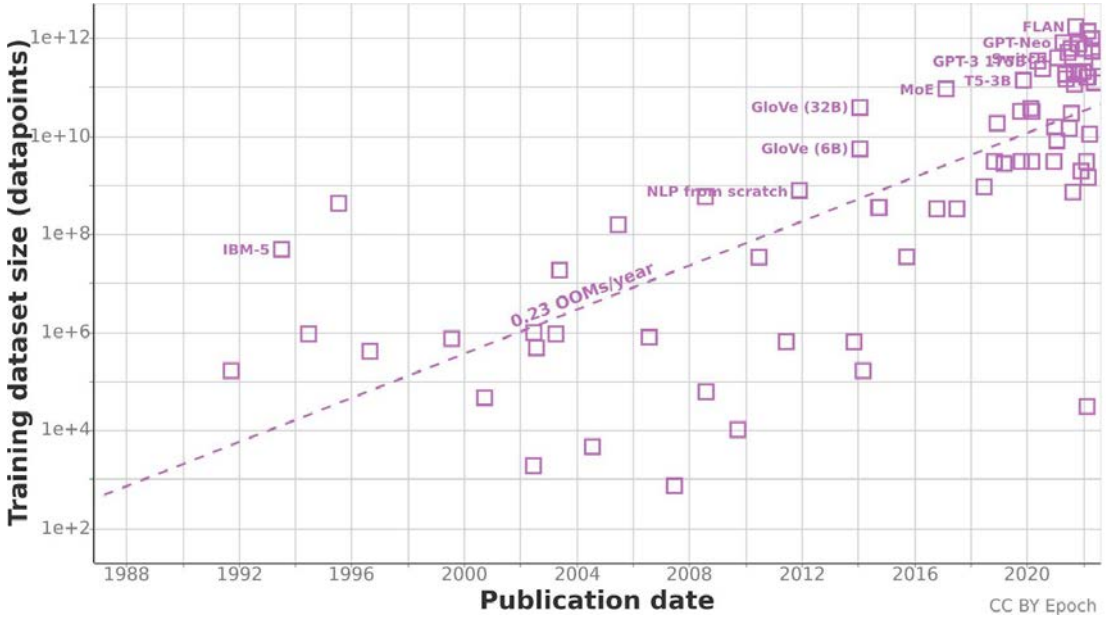
Books1 & Books2	
Romance	26.1%
Fantasy	13.6%
Science Fiction	7.5%
New Adult	6.9%
Young Adult	6.8%
Thriller	5.9%
Mystery	5.6%
Vampires	5.4%
Horror	4.1%
Other	18.0%

English Wikipedia	
Biography	27.8%
Geography	17.7%
Culture & Arts	15.8%
History	9.9%
Biology, Health, Medicine	7.8%
Sports	6.5%
Business	4.8%
Other society	4.4%
Science & Math	3.5%
Education	1.8%

More and more data

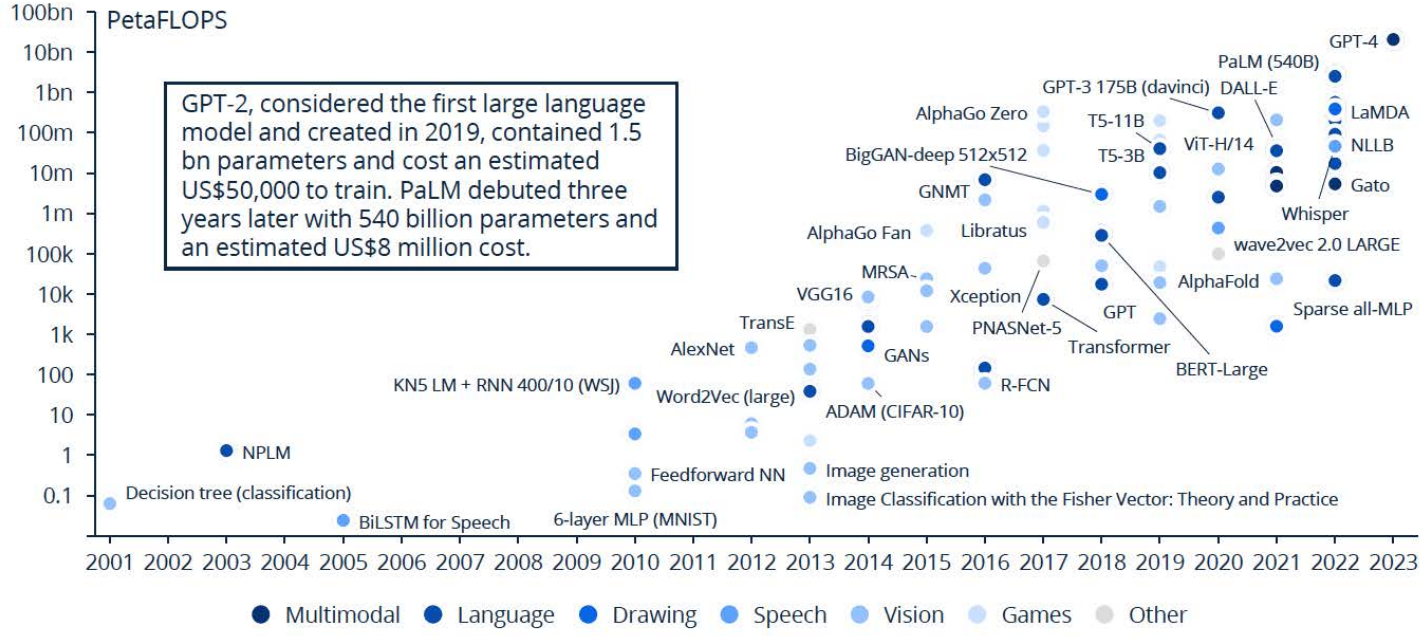
As a rule of thumb, to develop an efficient AI model, the number of training datasets required should be 10x more than each model parameter, also called degrees of freedom.

The '10' times rules aim to limit the variability and increase the diversity of data.

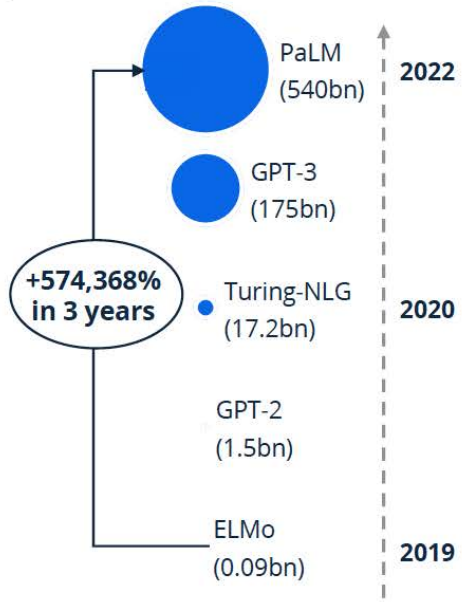


Serve tanta capacità di calcolo

Computation used to train notable artificial intelligence systems

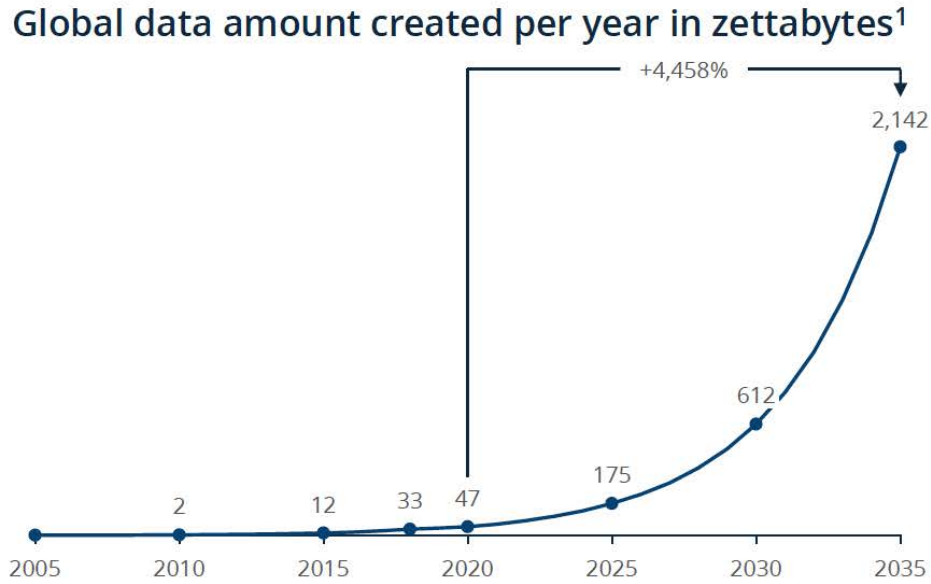


Language Model size in billions of parameters



I dati sono molto concentrati

Ma i dati sono molto concentrati alla fonte…



facebook …generates 4 petabytes of data per day, all that data is stored in what is known as the Hive, which contains about 300 petabytes of data

WhatsApp …users (2.5 billion globally) exchange up to 65 billion messages daily

Google … receives over 40,000 search queries per second that amount to 3.5 billion searches daily and 1.2 trillion searches per year (2020)

amazon …hosts its estimated 1,000,000,000 gigabytes of data across more than 1,400,000 servers

I dati sono molto concentrati

Ma anche il loro prodotto…

Profiling and targeting data repositories overview for major tech players

Companies: NETFLIX, Spotify, facebook, Tencent 腾讯, PayPal, Microsoft, Alibaba.com, Google, Apple, amazon

gathering…

Customer Profiles

comprising…

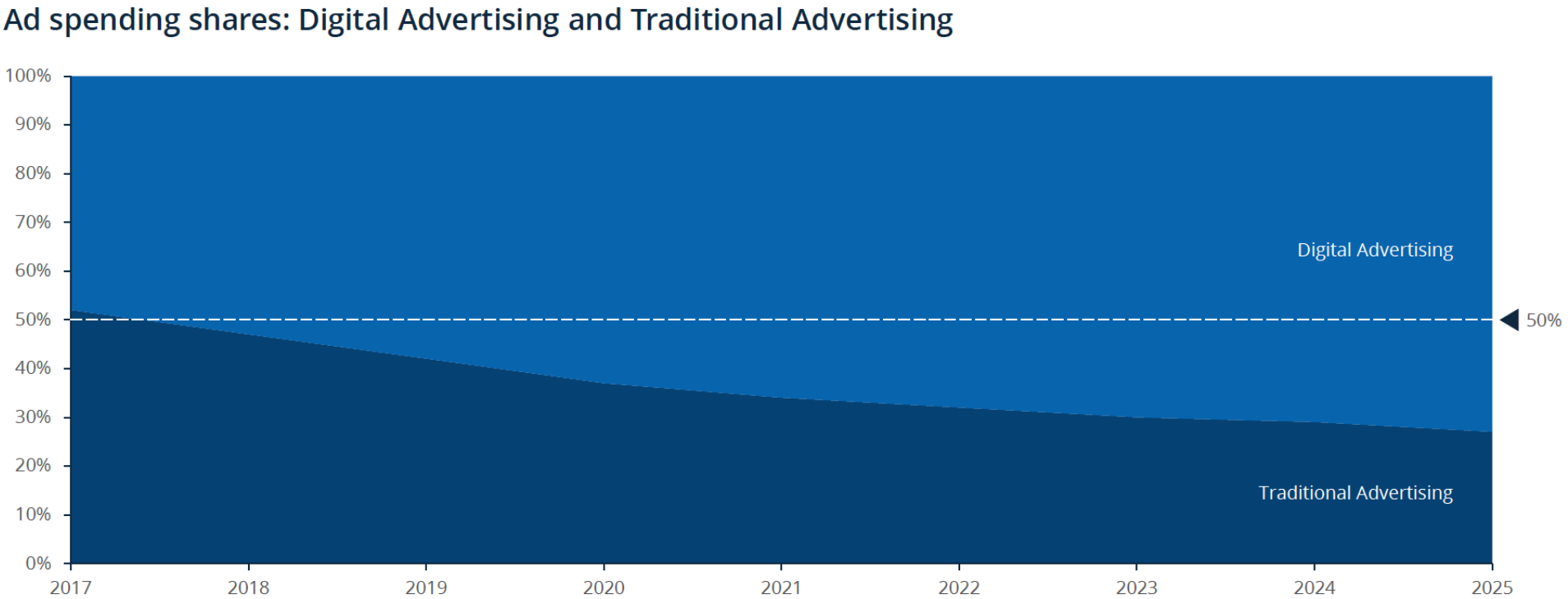
- **Demographics**: name, gender, birthday, relationship status, ethnicity, work status, education, income level
- **Health**: heart rate, weight, diseases, activity level, calorie intake, time slept
- **Devices**: devices in use, phone number, email Address, IP addresses, apps used, mails, contacts, phone calls, chat messages, calendar events, search history, websites visited, videos watched, music listened, browser information, video & photo uploads, purchase history, ads clicked, voice data, facial recognition data, fingerprint data
- **Geo**: current location, physical address, places visited (trip & vacation data), accommodations booked
- **Highly sensitive**: credit card information, bank data, dept data, net worth, government ids (social security or drivers license numbers)

enabling…

- Cross-domain and cross-device identity
- User profiles and segment creation
- Predictive marketing
- Advertising deployment
- Product optimization
- Fraud detection

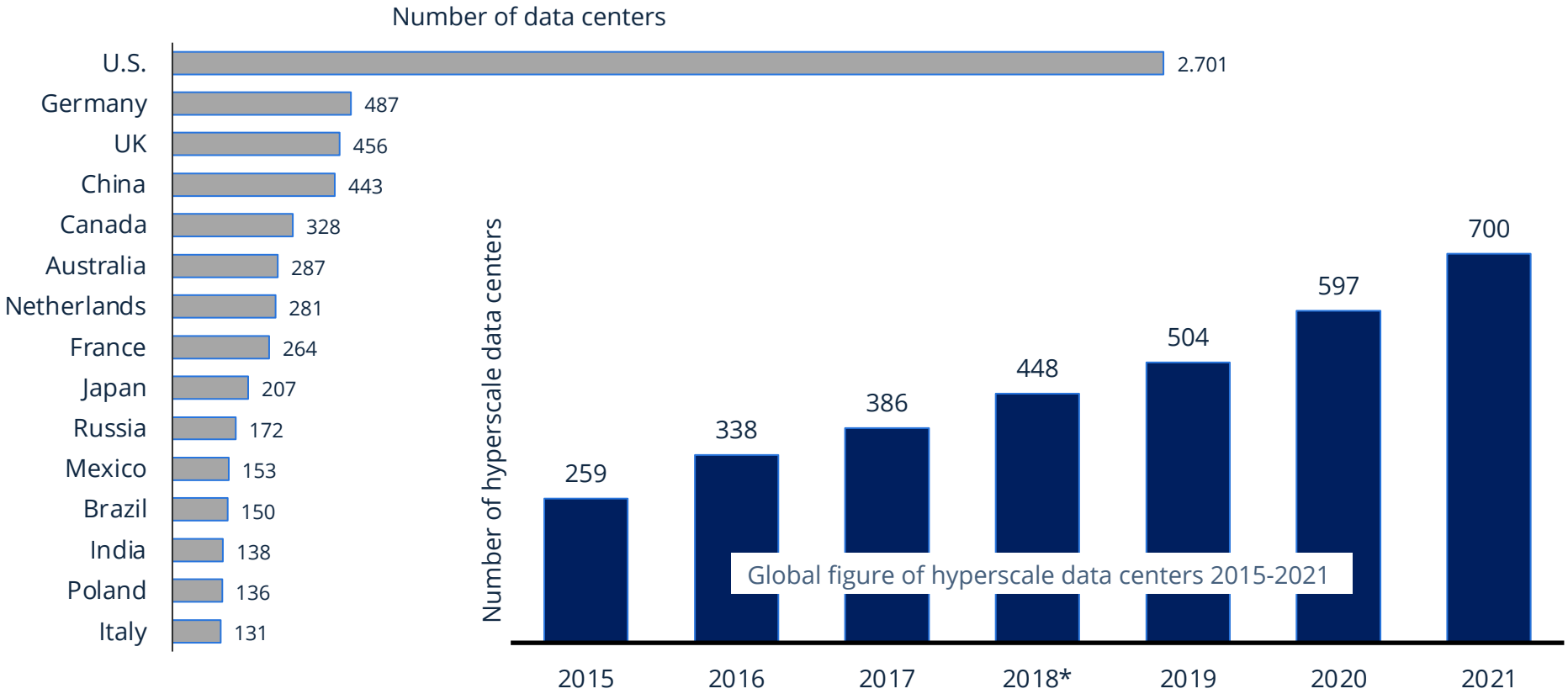
I dati sono molto concentrati

E anche il loro risultato (ad esempio, nell'advertising)…



La capacità di calcolo è concentrata

I data center sono l'habitat dei big data e della AI

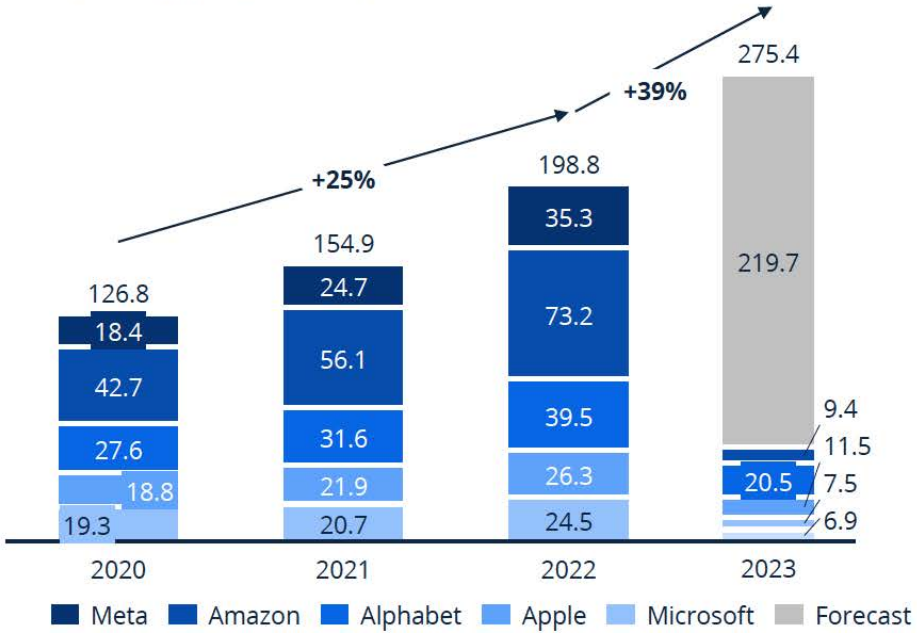


Source: Cloudscene, 2022

Source: Synergy Research Group, 2022

La spesa in R&D è esplosa tra i grandi del cloud…

R&D spending by big tech companies in billion US$(1)



Higher R&D spending than **Germany** in 2022 (143.1 bn.)

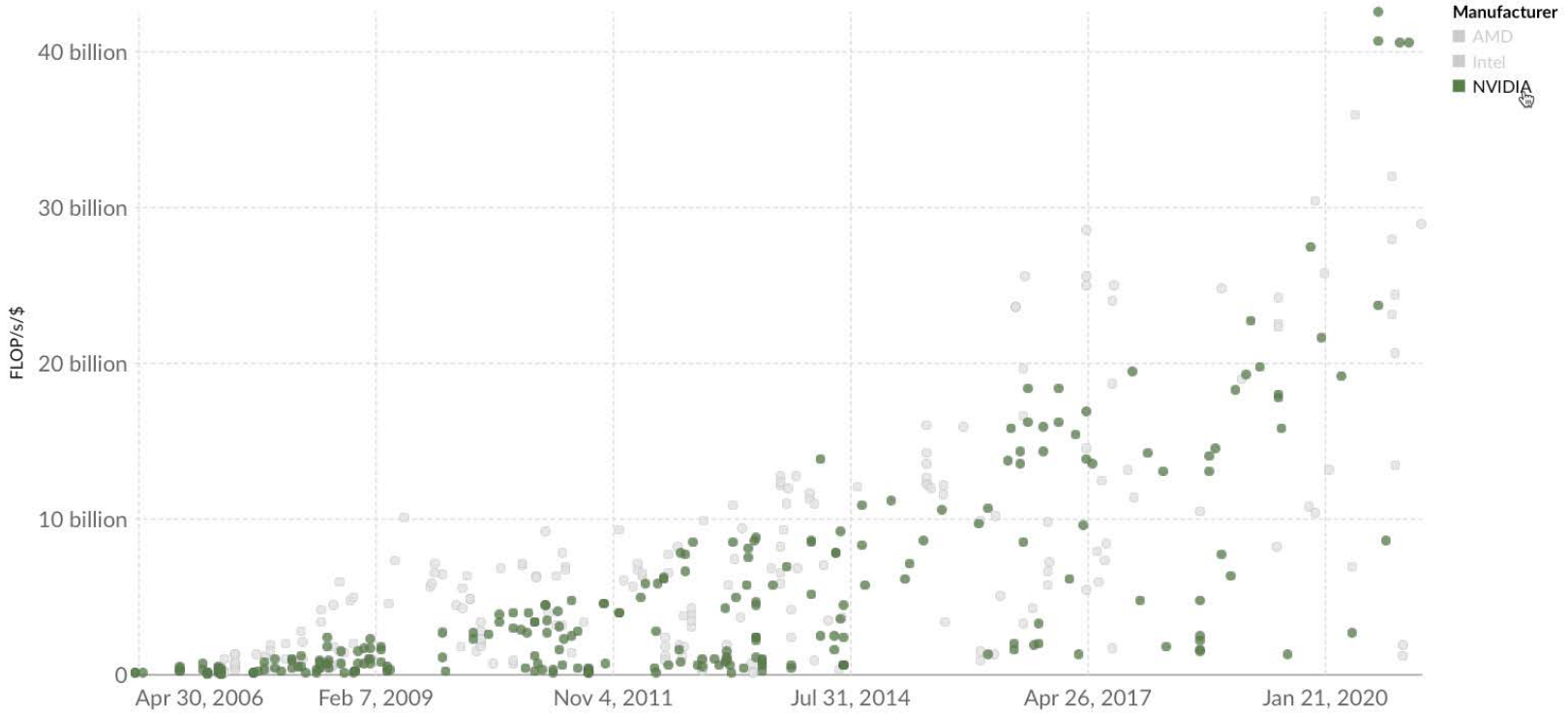
Higher R&D spending than **automotive** in 2022 (124 bn.)

Higher R&D spending than **GDP in Hungary** in 2022 (198 bn.)

- Microsoft Expanded partnership with OpenAI, launched Bing AI and Microsoft 365 Copilot, reported 2,500 Azure OpenAI customers, and expects AI to drive cloud growth.
- Alphabet Combined DeepMind and Google Brain, launched chatbot Bard, showcased AI-focused search functions, and aims to boost cloud revenues with AI.
- amazon Investing in generative AI and LLMs, launched AWS Bedrock for LLMs, and will use AI to drive advertising business growth.
- Meta Focused on AI for R&D and Metaverse topics, monetized content with AI recommendations, and leverages AI for business messaging and WhatsApp profitization.
- Apple Focused previously on hardware, expanded services segment, uses AI in products but needs to innovate in order to maintain market advantage.

Nvidia sta prendendo un ruolo guida nell'AI

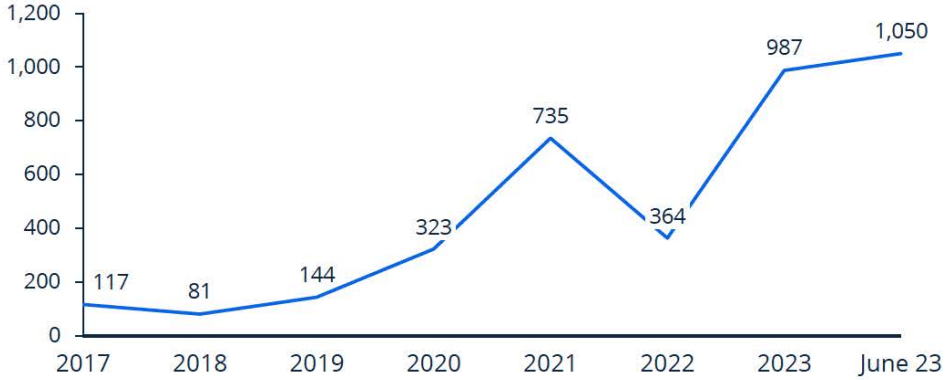
Per le sue performance…



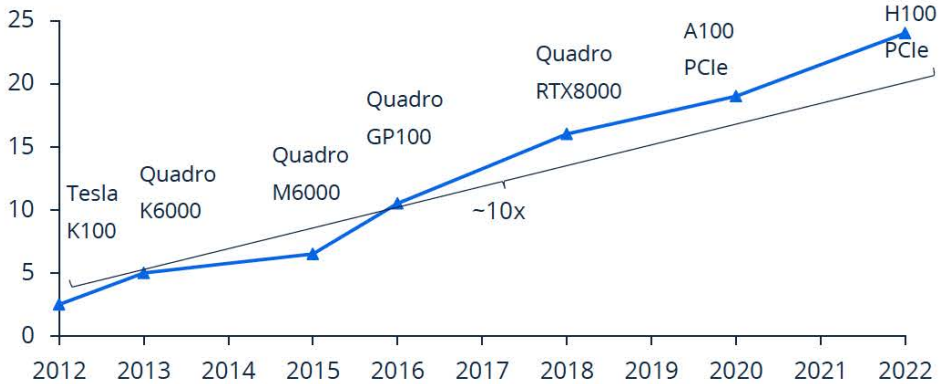
Nvidia sta prendendo un ruolo guida nell'AI

Anche grazie a partnership e collaborazioni

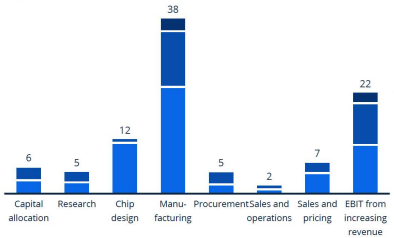
Nvidia's company valuation in billion USD



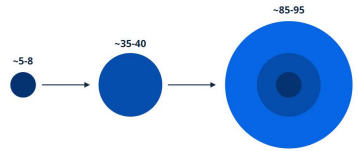
Nvidia GPU Performance measured in TFLOPS(1)



Impact on earnings EBIT by semiconductor key activities in US$ billion



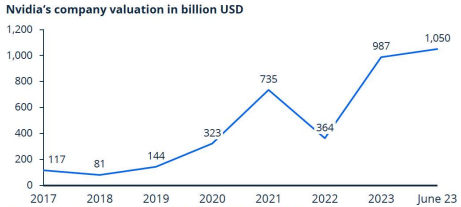
Impact of artificial intelligence on semiconductors EBIT in US$ billion

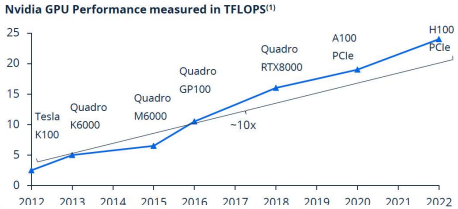


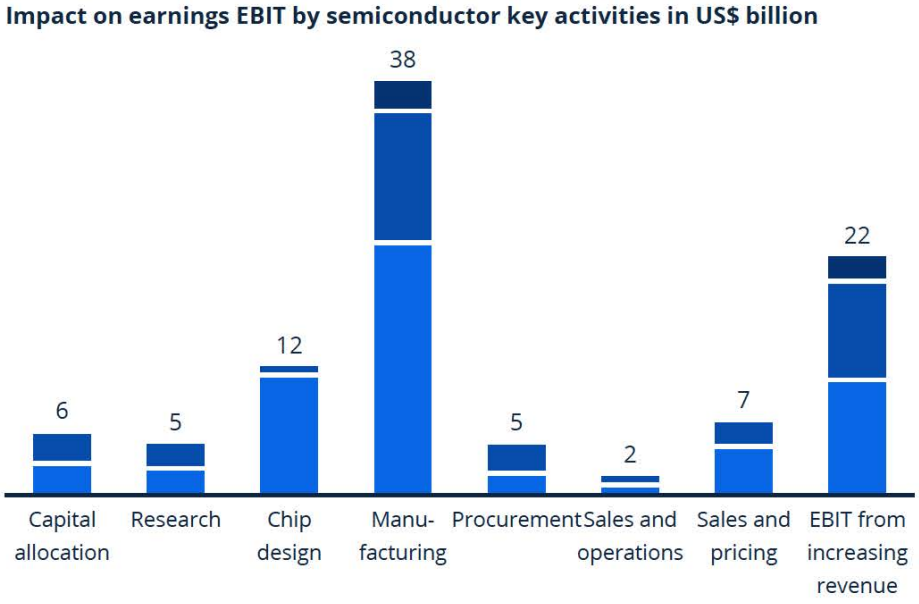
Current gains | Near-term potential(1) | Long-term potential(2)

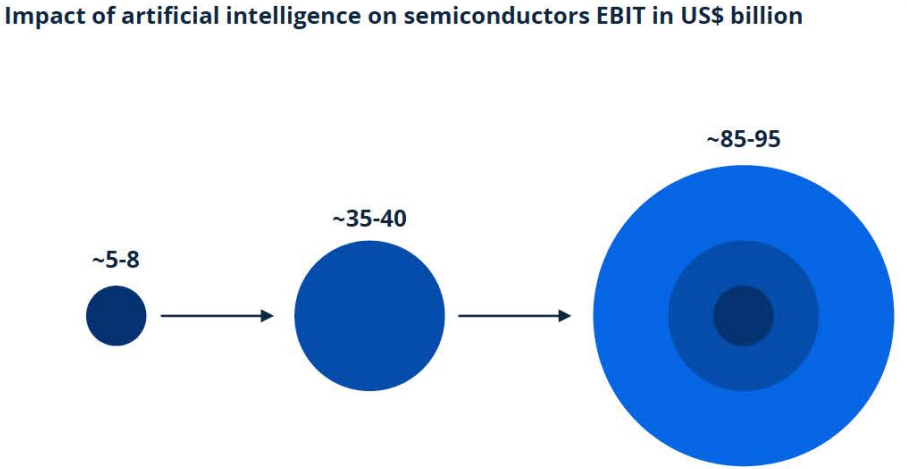
Nvidia sta prendendo un ruolo guida nell'AI

Anche grazie a partnership e collaborazioni



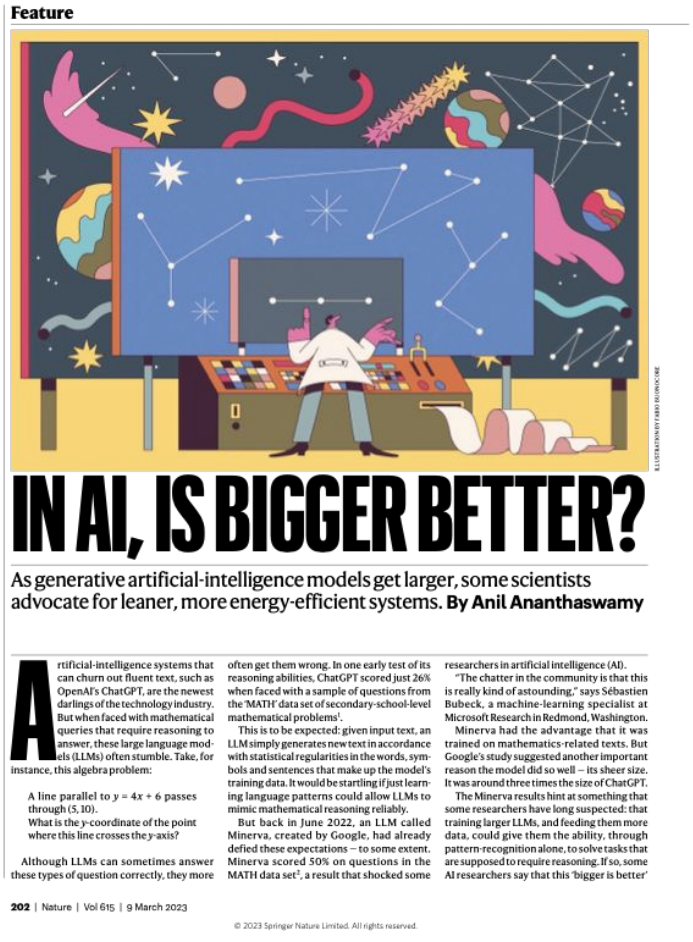






I LLM sono molto energivori

"Google reported that training PaLM took about 3.4 gigawatt-hours over about two months. That's the annual energy consumption of about 300 US households."



Feature

IN AI, IS BIGGER BETTER?

As generative artificial-intelligence models get larger, some scientists advocate for leaner, more energy-efficient systems. **By Anil Ananthaswamy**

Artificial-intelligence systems that can churn out fluent text, such as OpenAI's ChatGPT, are the newest darlings of the technology industry. But when faced with mathematical queries that require reasoning to answer, these large language models (LLMs) often stumble. Take, for instance, this algebra problem:

> A line parallel to $y = 4x + 6$ passes through (5, 10).
> What is the y-coordinate of the point where this line crosses the y-axis?

Although LLMs can sometimes answer these types of question correctly, they more often get them wrong. In one early test of its reasoning abilities, ChatGPT scored just 26% when faced with a sample of questions from the 'MATH' data set of secondary-school-level mathematical problems[1].

This is to be expected: given input text, an LLM simply generates new text in accordance with statistical regularities in the words, symbols and sentences that make up the model's training data. It would be startling if just learning language patterns could allow LLMs to mimic mathematical reasoning reliably.

But back in June 2022, an LLM called Minerva, created by Google, had already defied these expectations — to some extent. Minerva scored 50% on questions in the MATH data set[2], a result that shocked some researchers in artificial intelligence (AI).

"The chatter in the community is that this is really kind of astounding," says Sébastien Bubeck, a machine-learning specialist at Microsoft Research in Redmond, Washington.

Minerva had the advantage that it was trained on mathematics-related texts. But Google's study suggested another important reason the model did so well — its sheer size. It was around three times the size of ChatGPT.

The Minerva results hint at something that some researchers have long suspected: that training larger LLMs, and feeding them more data, could give them the ability, through pattern-recognition alone, to solve tasks that are supposed to require reasoning. If so, some AI researchers say that this 'bigger is better'

I LLM sono molto energivori

The New York Times

SUBSCRIBER-ONLY NEWSLETTER

Climate Forward

A.I. Could Soon Need as Much Electricity as an Entire Country

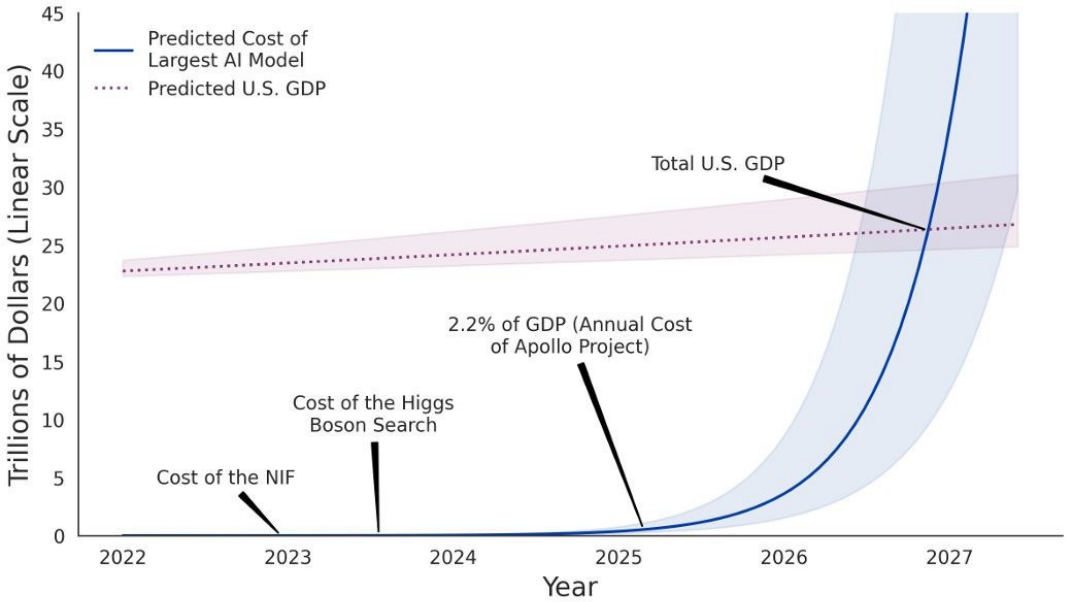
Behind the scenes, the technology relies on thousands of specialized computer chips.

"…by 2027 A.I. servers could use between 85 to 134 terawatt hours (Twh) annually"

More and more money

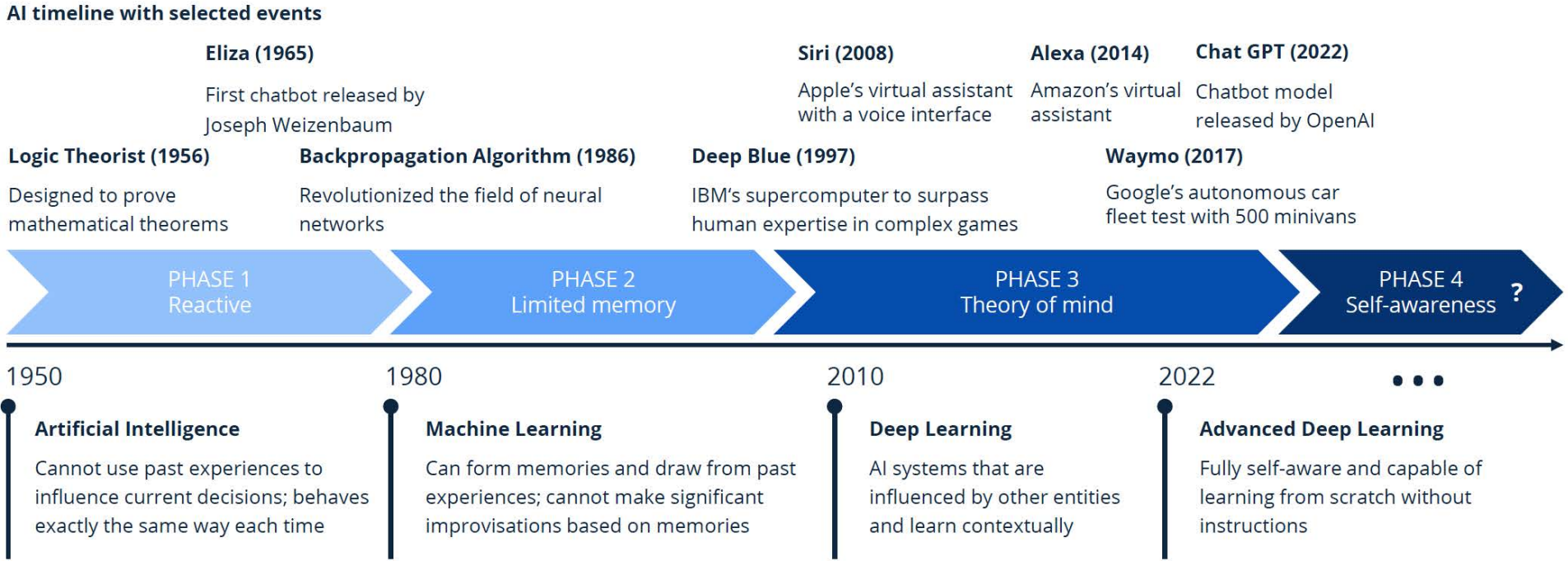
OpenAI has projected that the cost of training large AI models will increase from $100 million to $500 million by 2030, with the cost of training a single model ranging from $3 million to $12 million.

The cost of training a model on a large dataset can be even higher, reaching up to $30 million

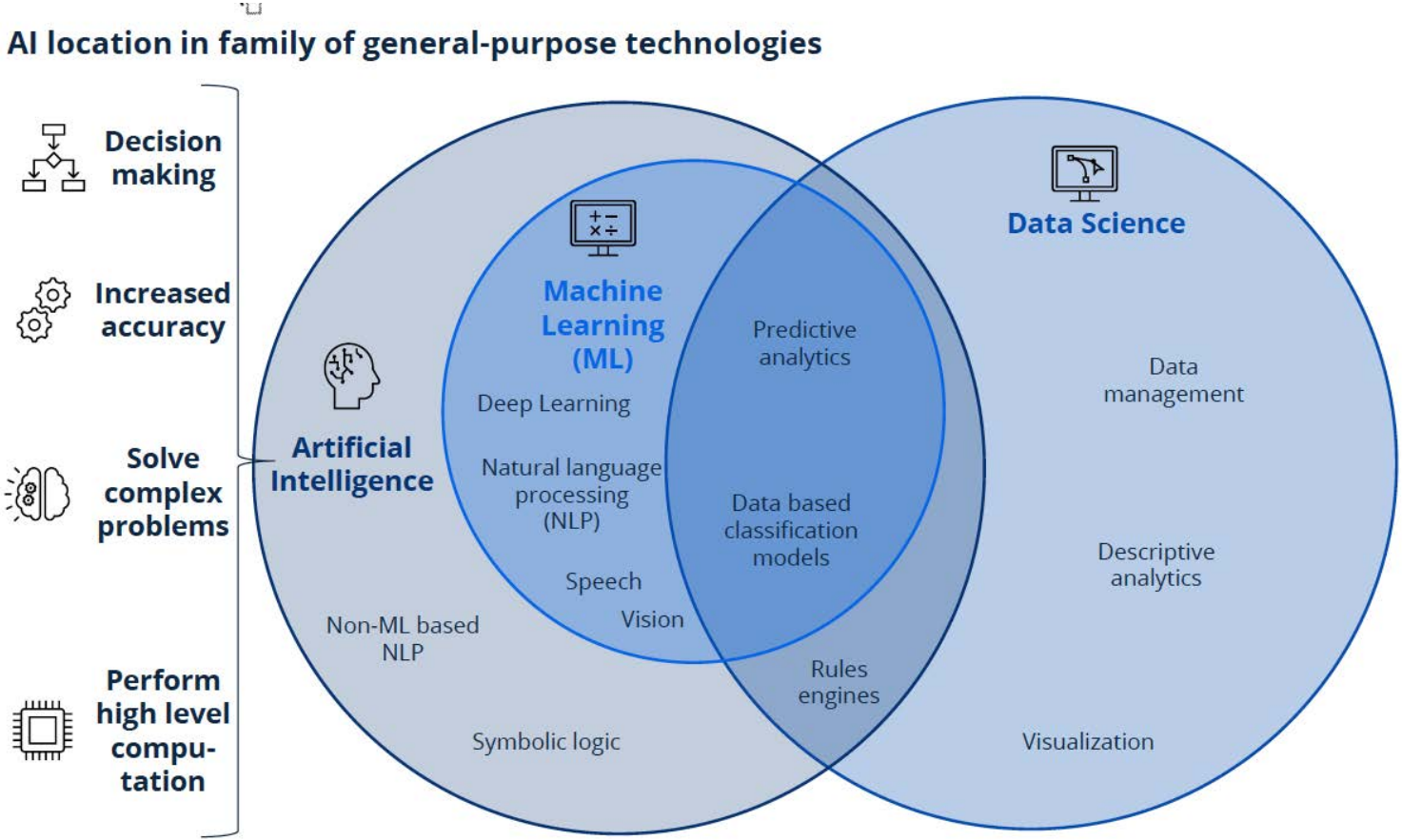


Source: CSET. Note: The blue line represents growing costs assuming compute per dollar doubles every four years, with error shading representing no change in compute costs or a doubling time as fast as every two years. The red line represents expected GDP at a growth of 3 percent per year from 2019 levels with error shading representing growth between 2 and 5 percent.

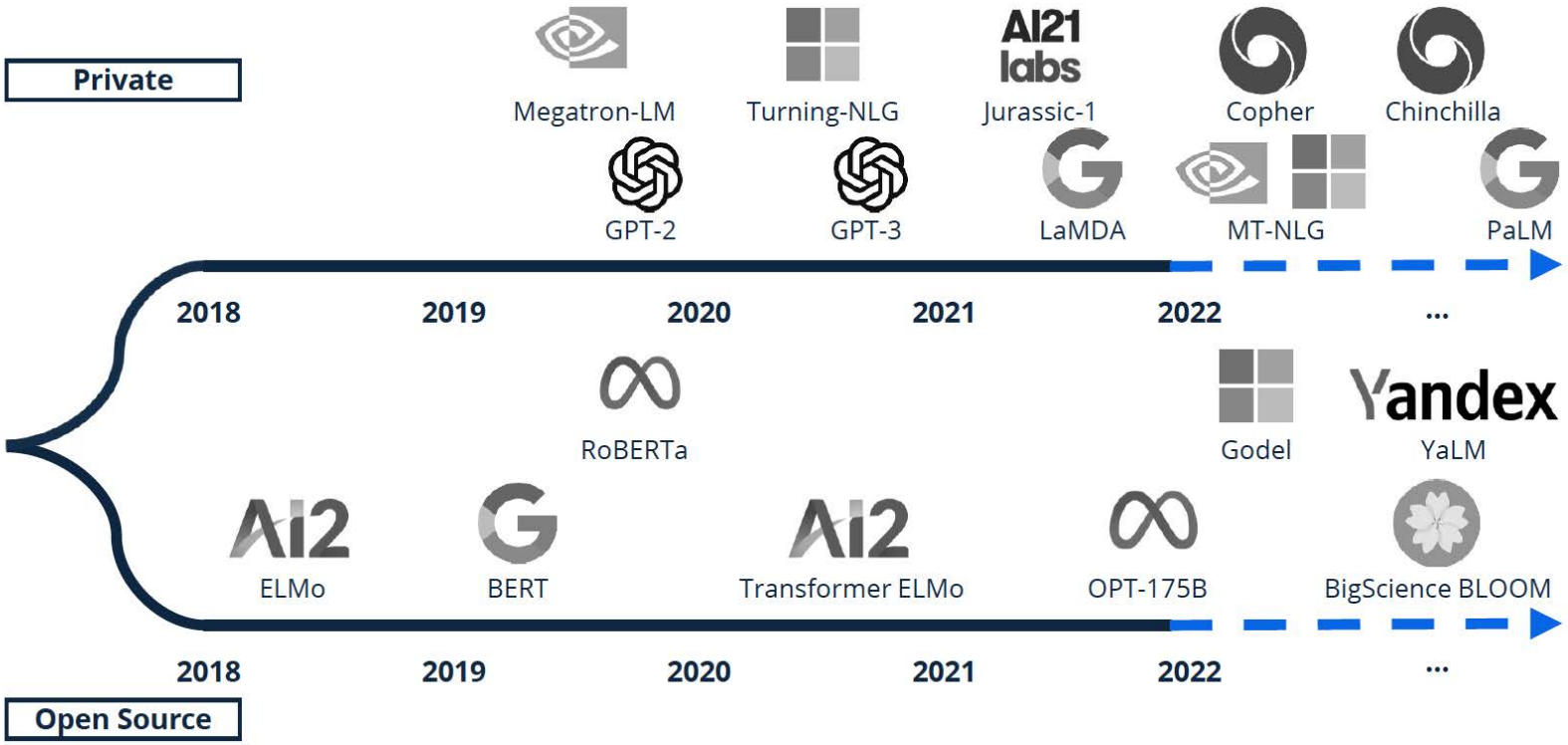
Una cronologia



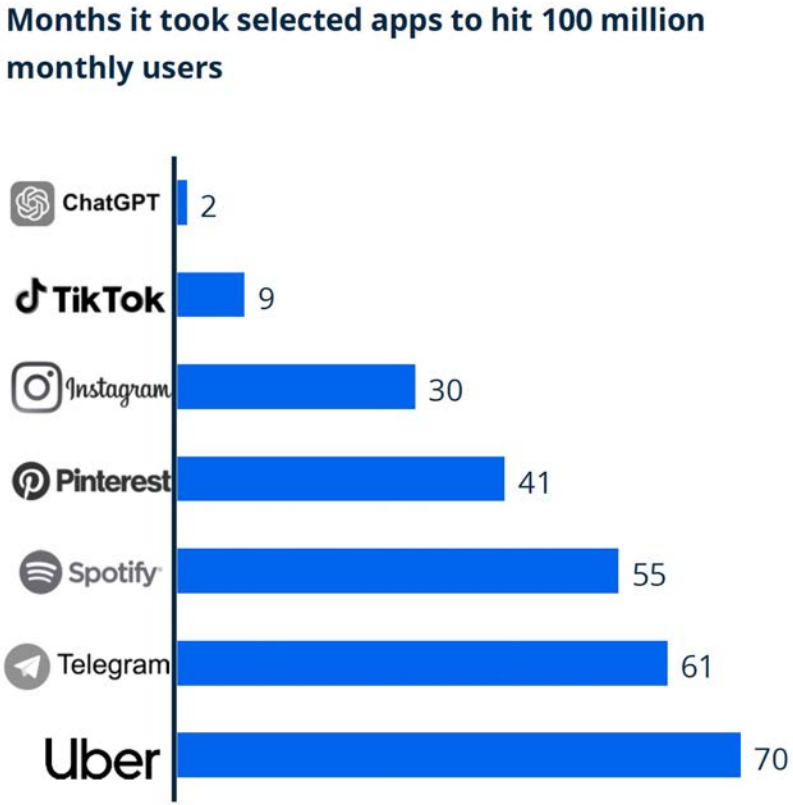
Una vista d'insieme

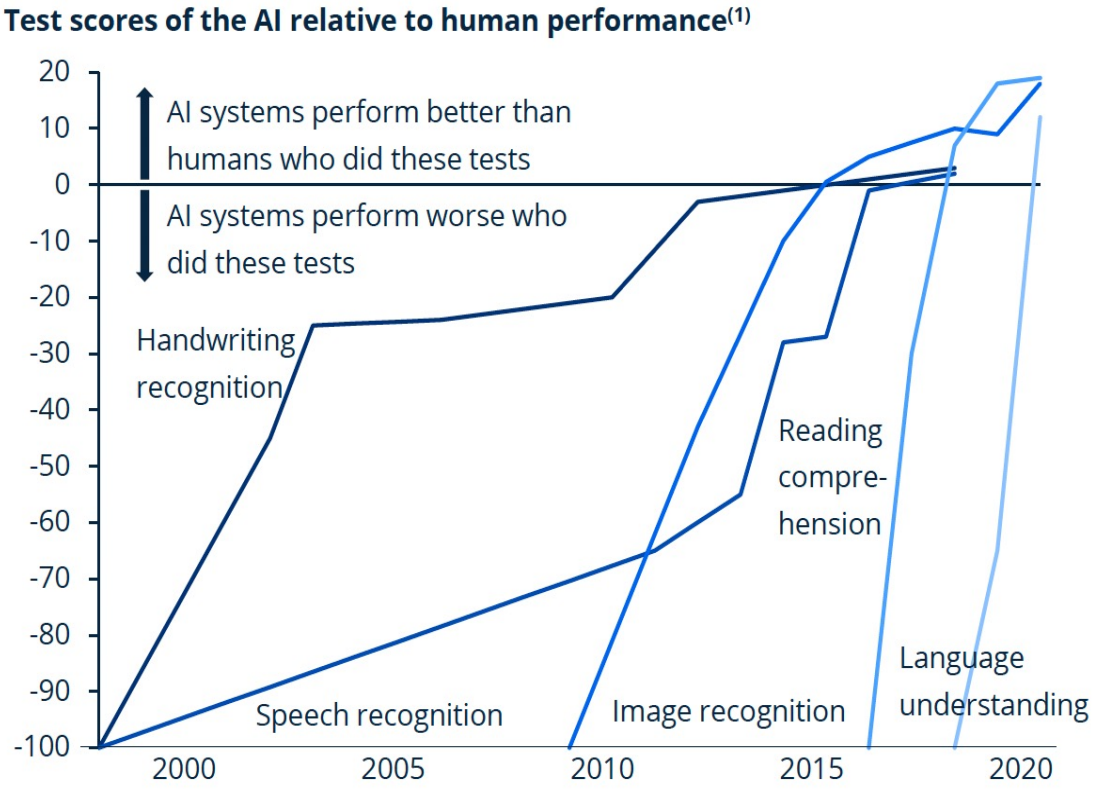


C’è stata recentemente un’esplosione di modelli

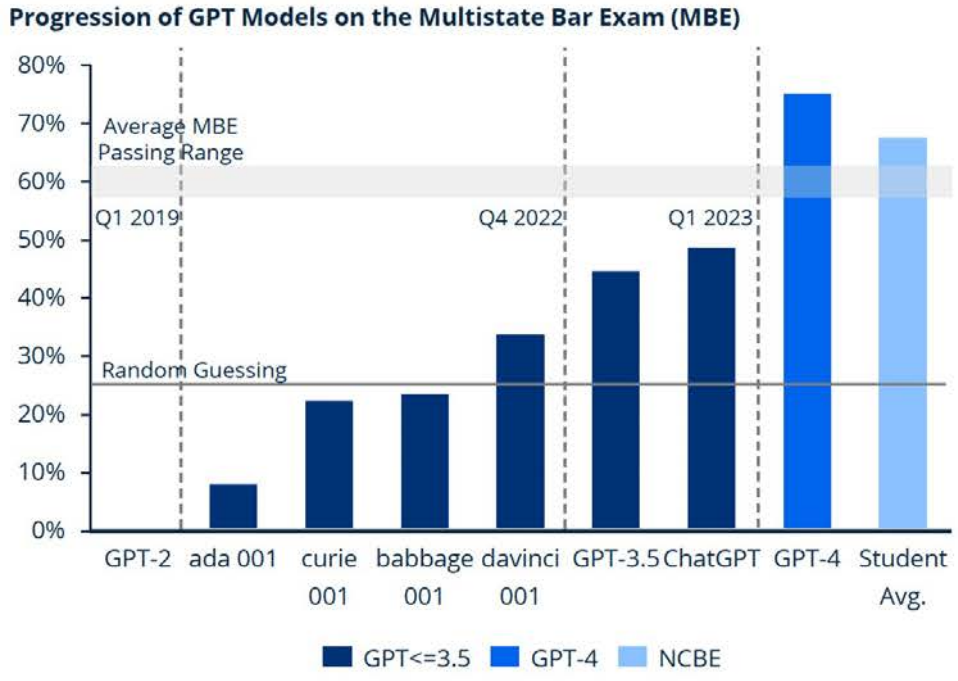


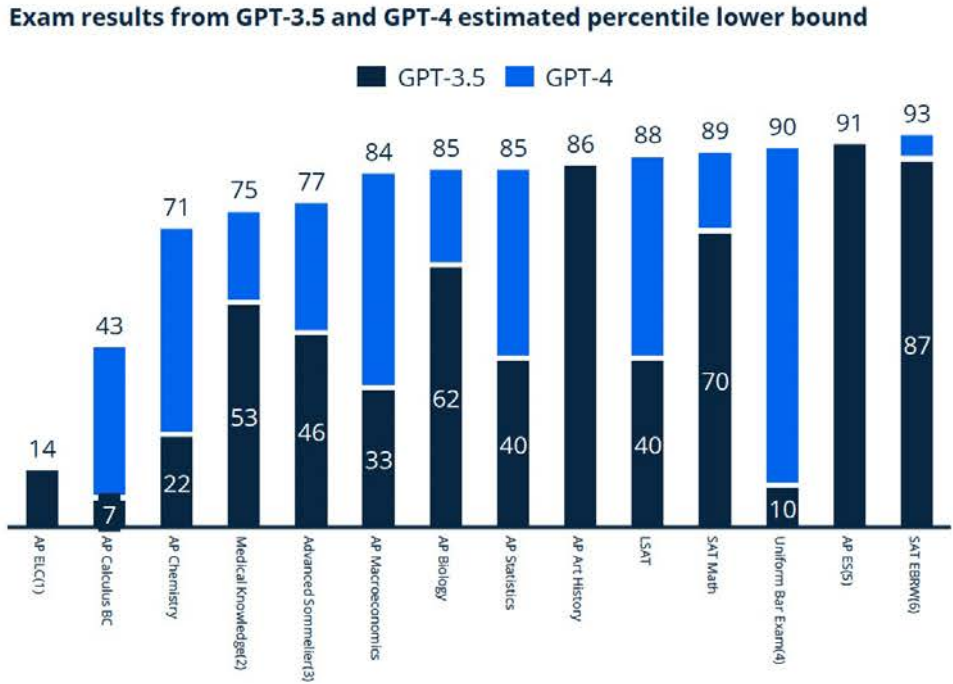
E un grande successo di pubblico e di risultati





E un grande successo di pubblico e di risultati





L'AI generativa può produrre un ampio range di risultati

Generative AI

Generating content and ideas	Improving efficiency	Personalizing experiences
Creating new, unique outputs across a range of modalities, such as a video advertisement or even a new protein with antimicrobial properties.	Accelerating manual or repetitive tasks, such as writing e-mails, coding, or summarizing large documents.	Creating content and information tailored to a specific audience, such as chatbots for personalized customer experiences or targeted advertisements based on specific customer behavior patterns.

Types

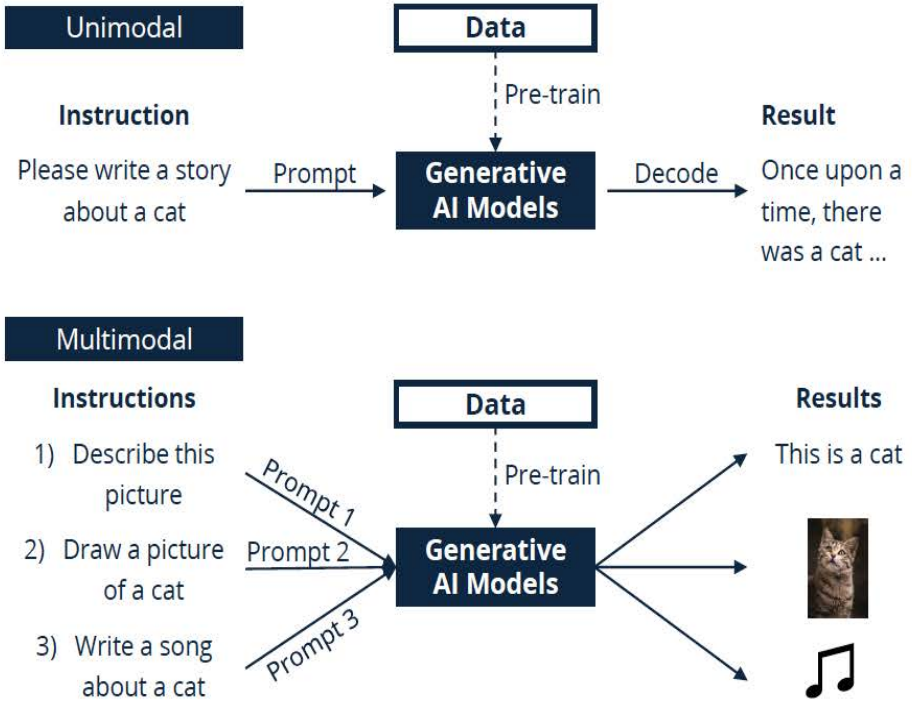
Image (video) generation	Text generation	Audio generation
Midjourney DallE	ChatGPT Hugging Face	Tacotron 2 WaveGAN

Examples of Generative AI technology

	Input	Output
Chat GPT	"Write a limerick about data taking over the world, include the word egg"	"Data taking over the world, it's true it's growing like an egg that's due, It's everywhere we go, it's taking over slow. But it's the future, there's nothing we can do"
Midjourney	prompt for image generation, such as rabbit wearing a hat	creating fictional and realistic pictures based on input prompts 
Codex	„Write a script to change a website's background color in Python"	Python code for changing the website's background color

L'AI generativa può produrre un ampio range di risultati

Unimodal and multimodal Generative AI models[1]

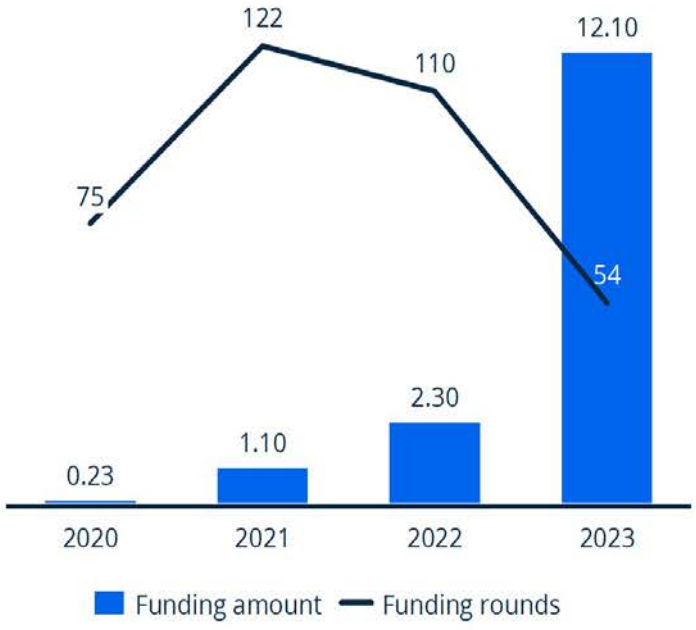


Technology used in Generative AI models

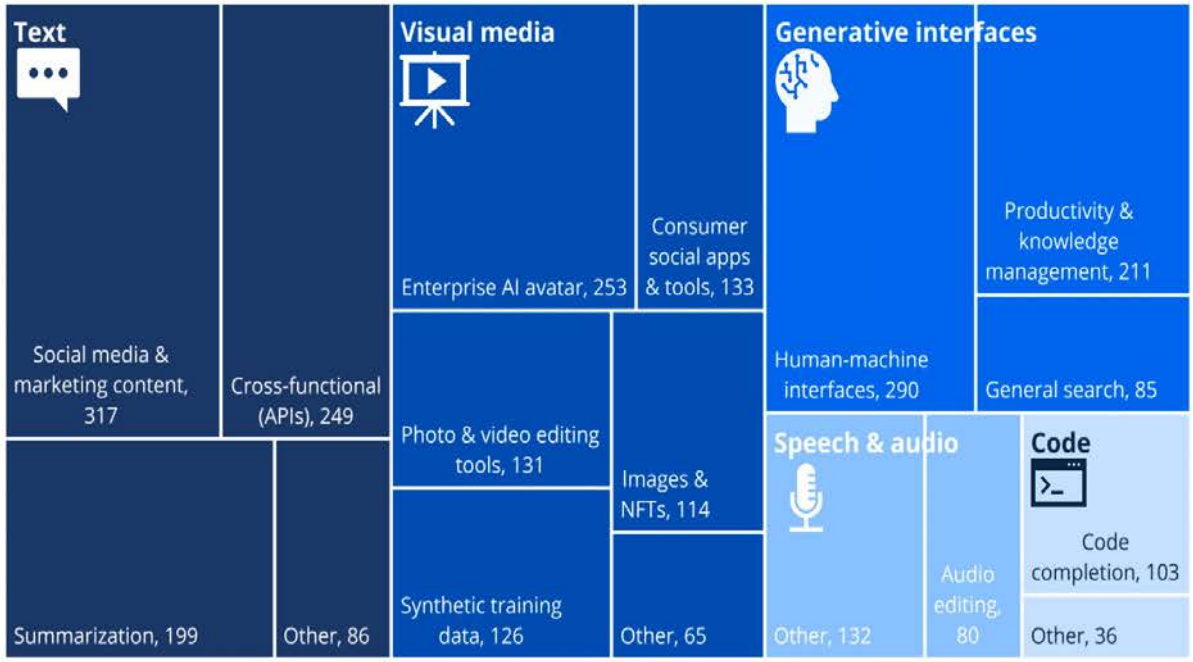
Generative Adversarial Networks **(GANs)**	Variational Autoencoders **(VAEs)**	**Flow-based Generative Models**
• GANs are highly effective in generating realistic multimedia, including deepfakes, movie dubs, and images from text descriptions. • They utilize a generator network to create content and a discriminator network to assess the realism of the generated data. • Through iterative training, GANs continuously improve their ability to generate data that closely resembles reality.	• VAEs are commonly used in signal analysis tasks. • They utilize encoder and decoder networks within an autoencoder framework. • VAEs focus on efficient data representation and reconstruction, minimizing errors between original and reconstructed data. • Used for cleaning data, predictive analysis, data compression, and reducing data dimensionality.	• Flow-based generative models are commonly used for generative AI tasks. • A flow-model architecture is used to map a simple distribution to a complex data distribution. • Flow-based models learn the mapping through invertible transformations, allowing efficient sampling and likelihood estimation. • Used for generating realistic images, text, and other types of data.

L'AI generativa può produrre un ampio range di risultati

Funding amount in billion US$ and number of funding rounds for Generative AI



Distribution of funding for Generative AI in million US$ for 2021 to 2022



Le tribù dell'intelligenza artificiale

Simbolisti: si concentrano sull'uso di simboli, regole e logica per rappresentare la conoscenza e dedurre nuove informazioni. I loro algoritmi preferiti sono i sistemi basati su regole e gli alberi decisionali.

Bayesiani: Questa tribù valuta la probabilità di risultati diversi per fare inferenze probabilistiche. Spesso usano metodi bayesiani, inclusi algoritmi come i modelli Naive Bayes o Markov.

Connessionisti: I connessionisti si ispirano al funzionamento del cervello umano e si concentrano sui modelli di apprendimento attraverso le strutture delle reti neurali, che sono composte da nodi (come i neuroni) collegati da bordi (come le sinapsi).

Evoluzionisti: utilizzano algoritmi che imitano l'evoluzione biologica, come gli algoritmi genetici, per generare una varietà di soluzioni e poi farle evolvere nel tempo in base alla loro idoneità per un determinato scopo.

Analogizers: questo gruppo si concentra sulla ricerca e l'ottimizzazione di funzioni basate sull'analisi di somiglianze o analogie. Spesso impiegano macchine a vettori di supporto (SVM) che trovano il confine ottimale (o iperpiano) tra diverse classi di dati.

Le tribù dell'intelligenza artificiale

Simbolisti: si concentrano sull'uso di simboli, regole e logica per rappresentare la conoscenza e dedurre nuove informazioni. I loro algoritmi preferiti sono i sistemi basati su regole e gli alberi decisionali.

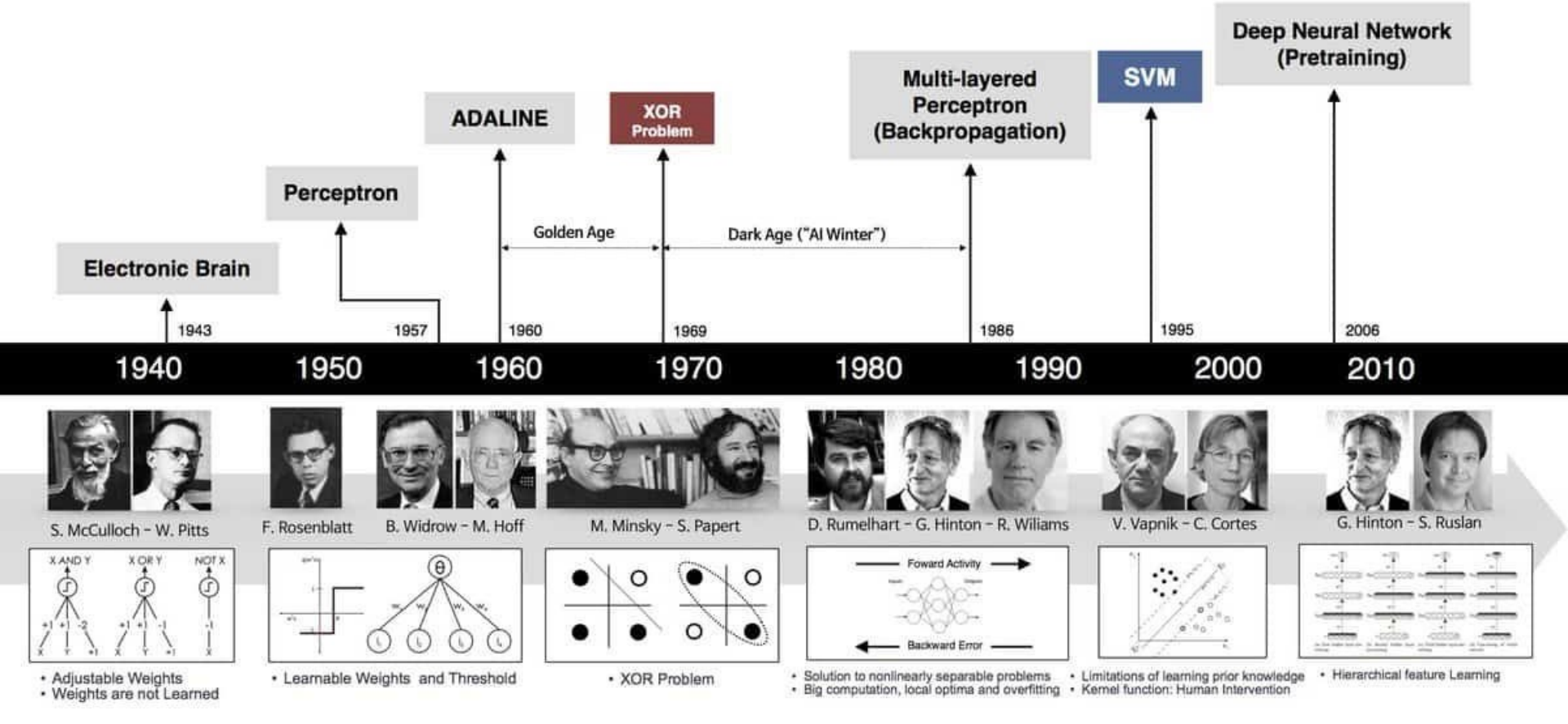
Bayesiani: Questa tribù valuta la probabilità di risultati diversi per fare inferenze probabilistiche. Spesso usano metodi bayesiani, inclusi algoritmi come i modelli Naive Bayes o Markov.

Connessionisti: I connessionisti si ispirano al funzionamento del cervello umano e si concentrano sui modelli di apprendimento attraverso le strutture delle reti neurali, che sono composte da nodi (come i neuroni) collegati da bordi (come le sinapsi).

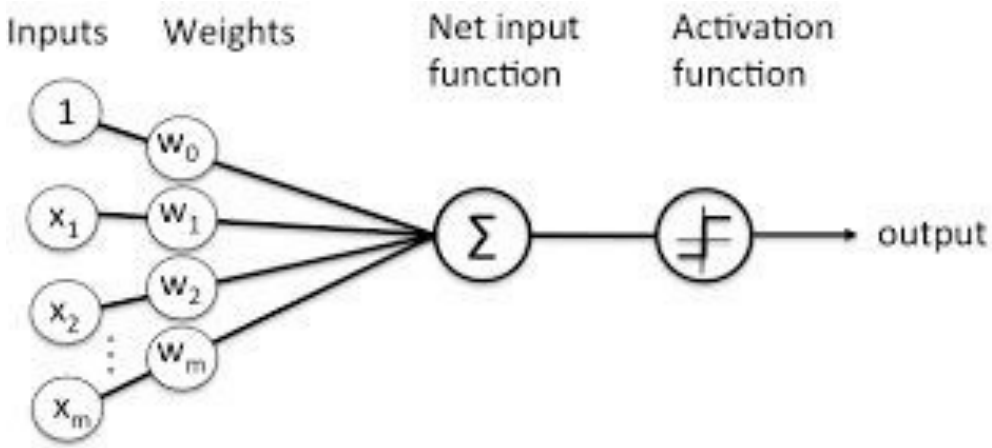
Evoluzionisti: utilizzano algoritmi che imitano l'evoluzione biologica, come gli algoritmi genetici, per generare una varietà di soluzioni e poi farle evolvere nel tempo in base alla loro idoneità per un determinato scopo.

Analogizers: questo gruppo si concentra sulla ricerca e l'ottimizzazione di funzioni basate sull'analisi di somiglianze o analogie. Spesso impiegano macchine a vettori di supporto (SVM) che trovano il confine ottimale (o iperpiano) tra diverse classi di dati.

I connessionisti

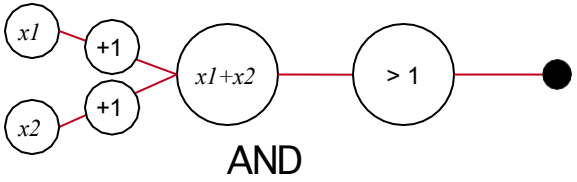


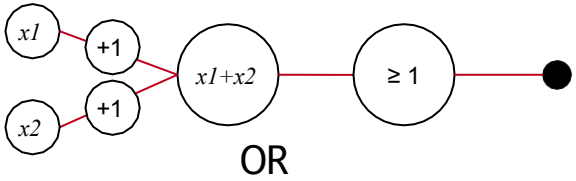
I neuroni artificiali

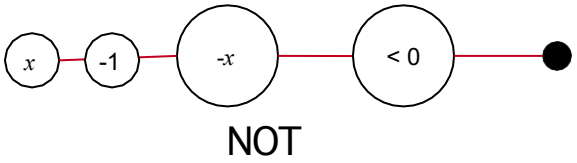


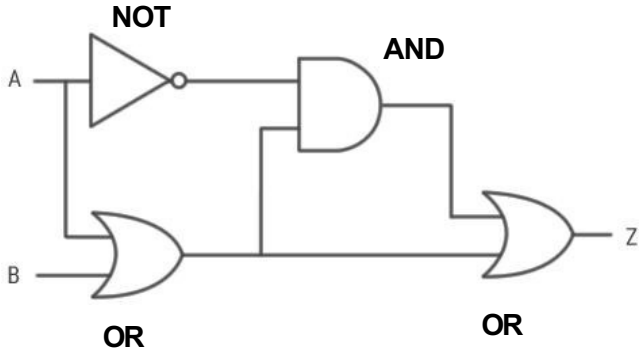
$$\sum_{i=1}^{n} w_i x_i + w_0 = 0$$

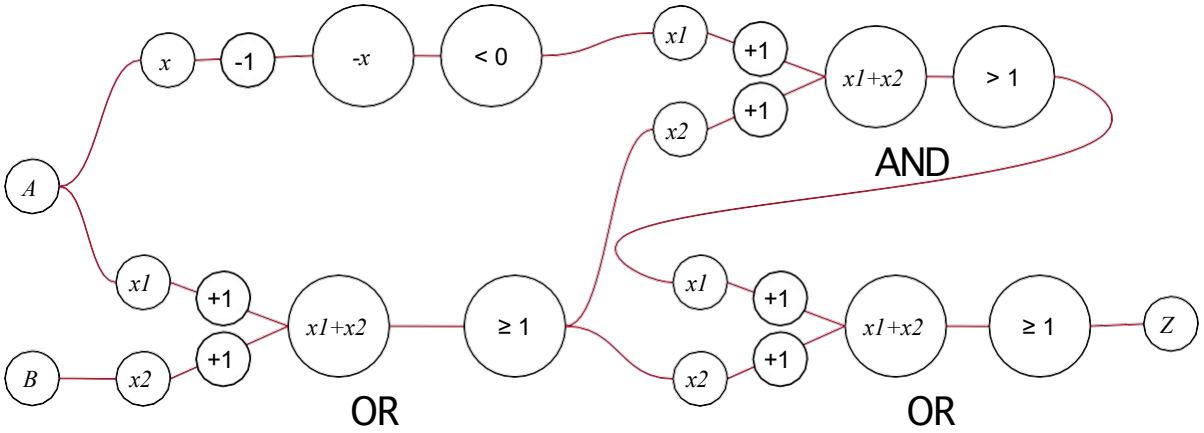
Logic interpretation







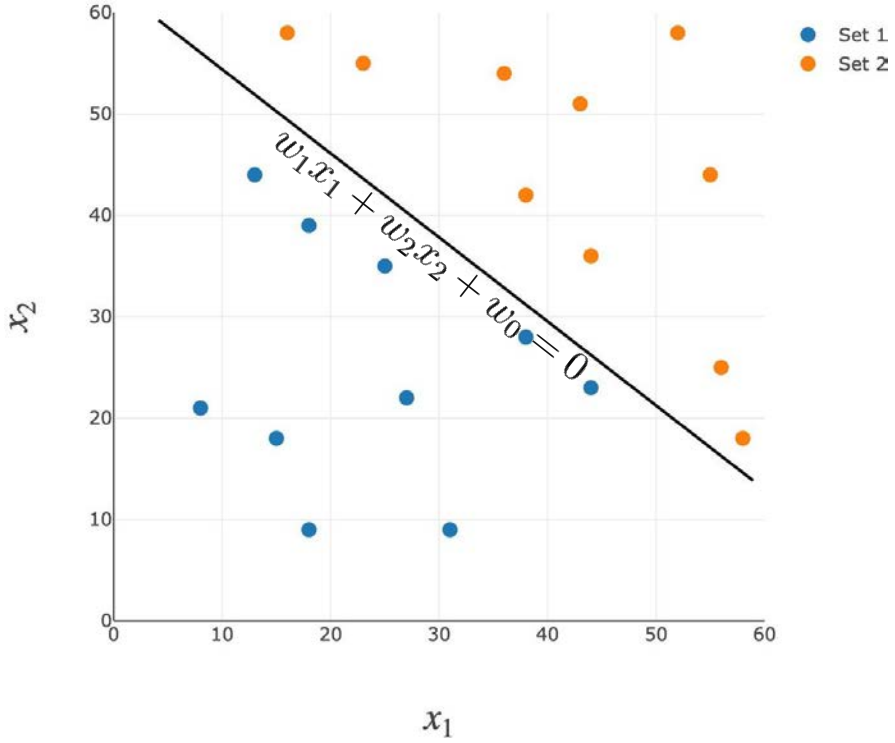


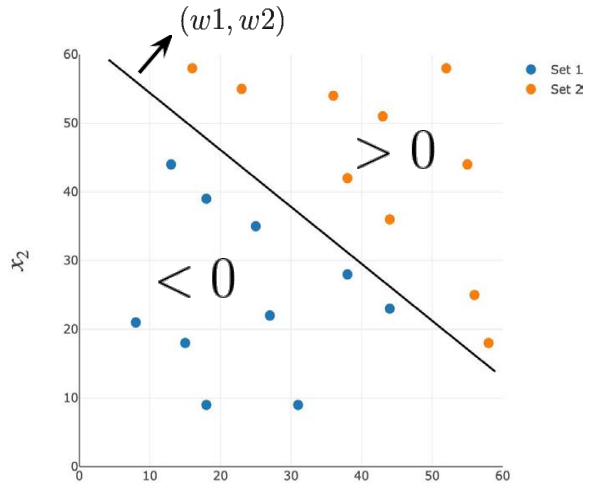


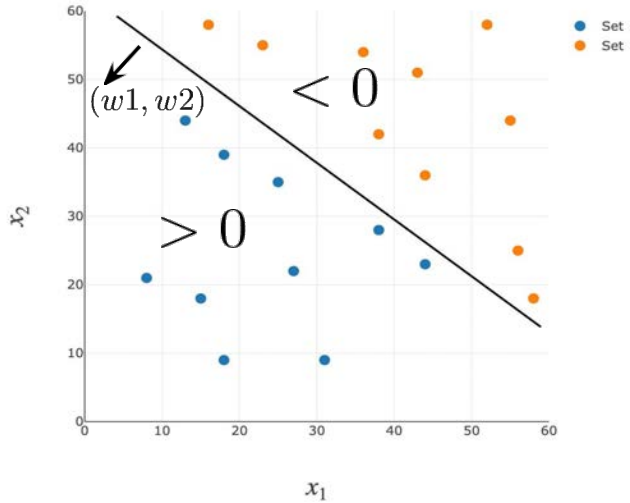
FUNCTIONAL COMPLETE!

La geometria dei neuroni

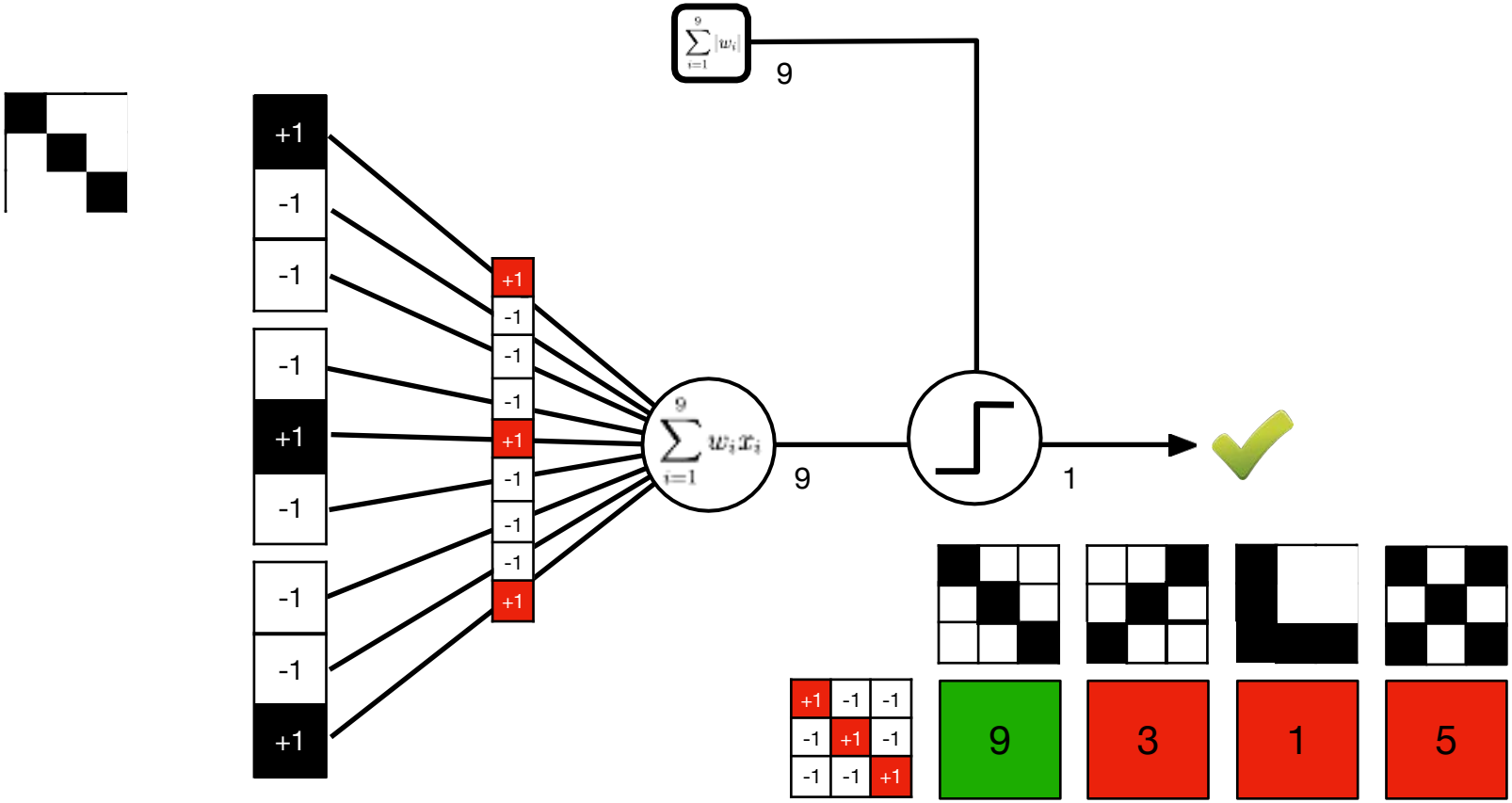
$$w_1 x_1 + w_2 x_2 + w_0 = 0$$



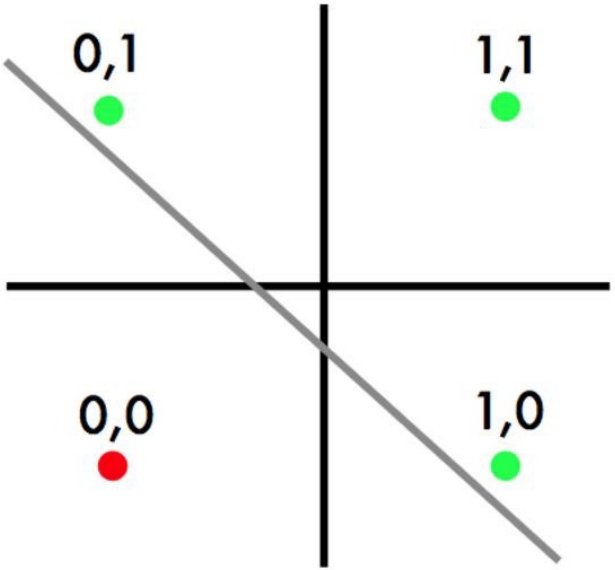




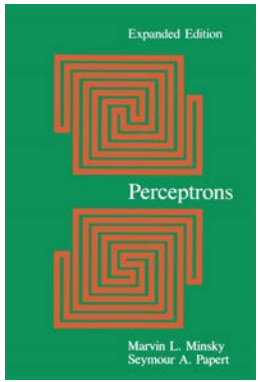
NEURON AS PATTERN MATCHING

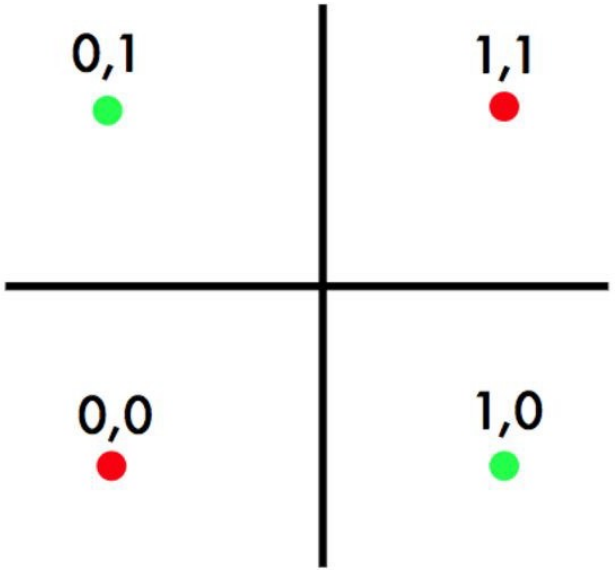


THE XOR (EQ) PROBLEM



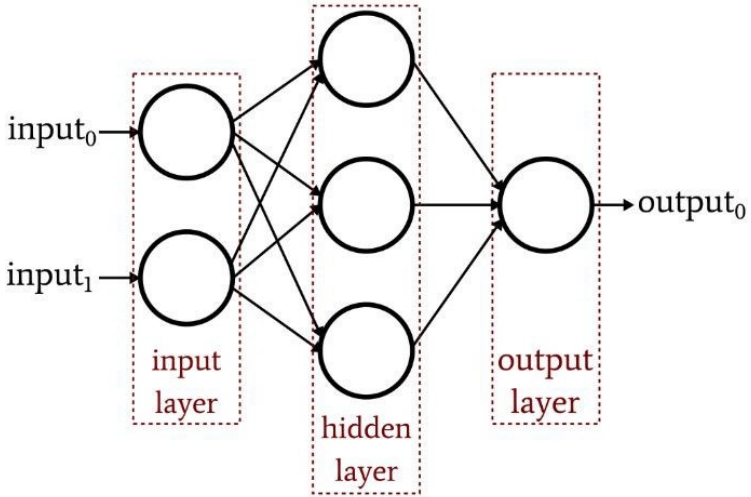
OR

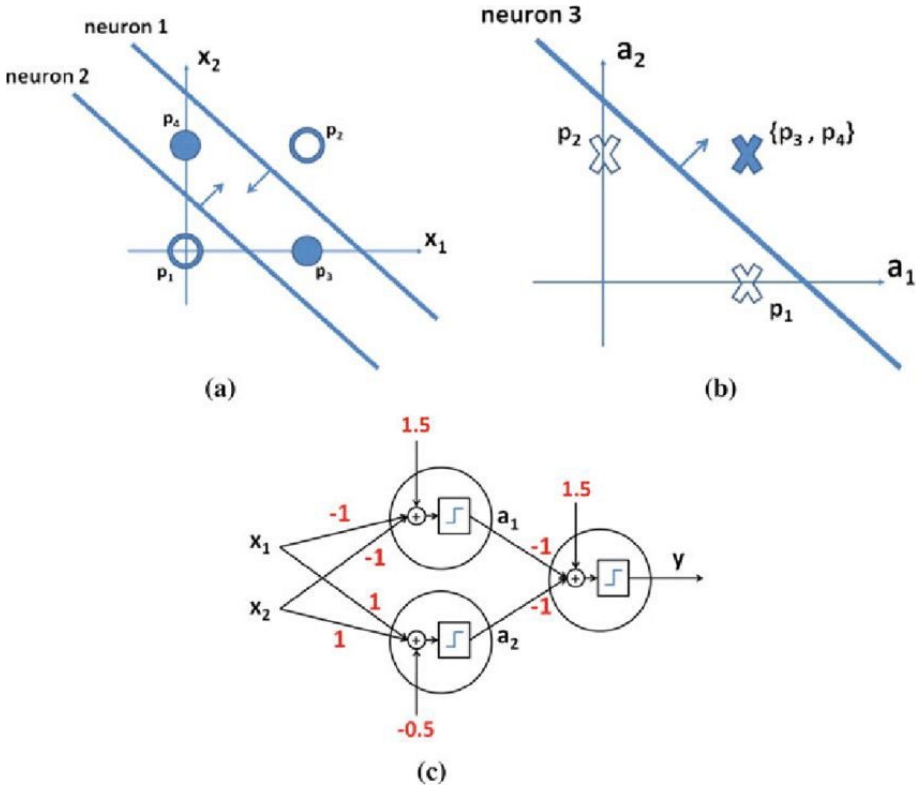


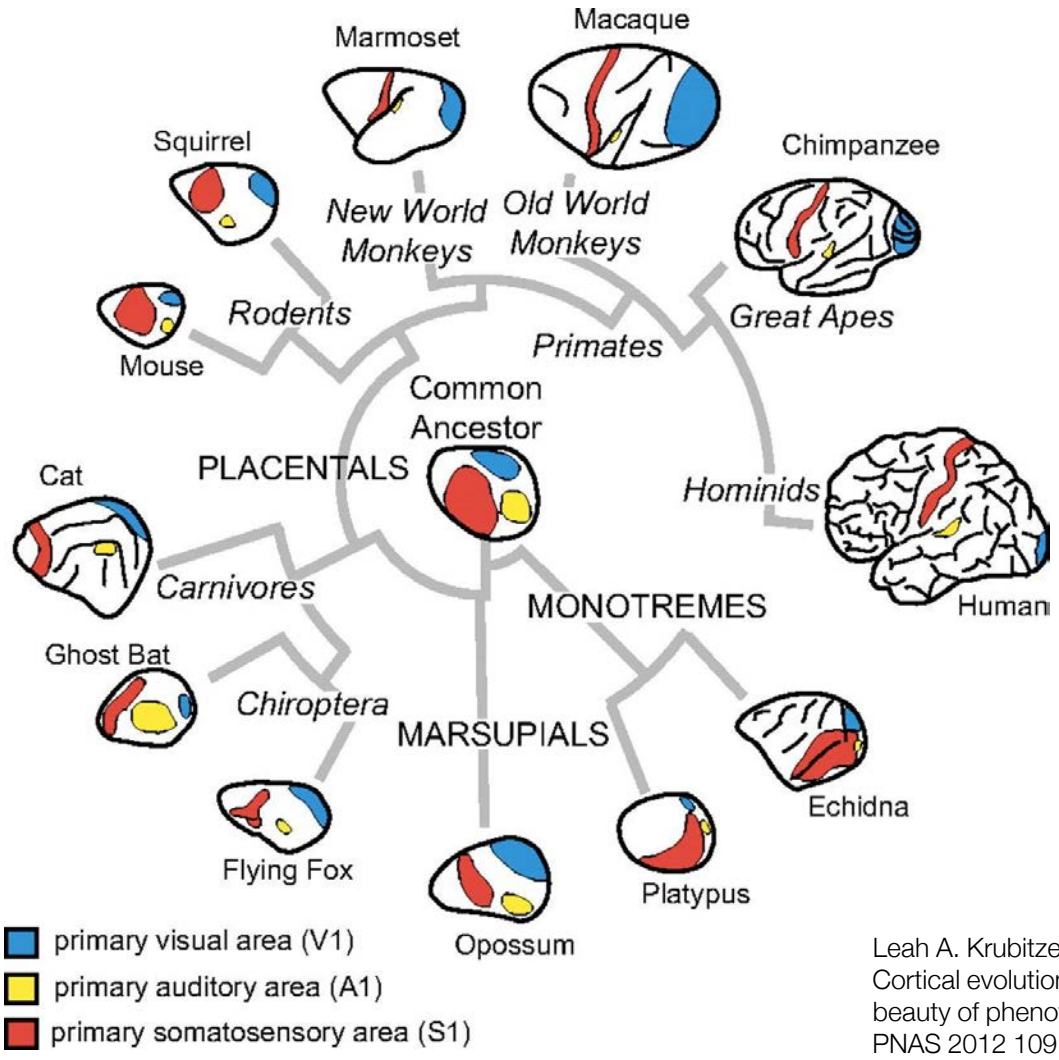


XOR

MULTILAYER PERCEPTRON



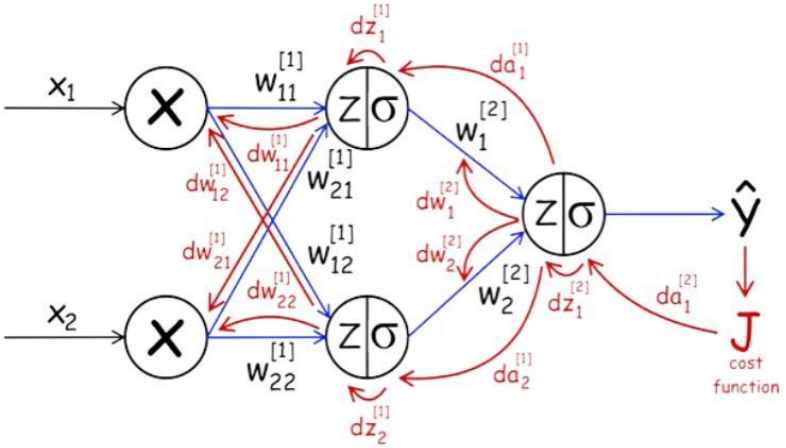




Leah A. Krubitzer and Adele M. H. Seelke
Cortical evolution in mammals: The bane and beauty of phenotypic variability
PNAS 2012 109 (Supplement 1) 10647-10654

The universal approximation theorem

A neural network with just one hidden layer can, given enough neurons and the right set of weights, approximate any continuous function.



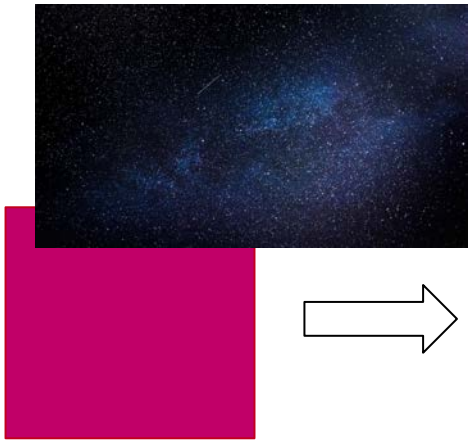
BACK-PROPAGATION CAN DO THAT

I binari dell'informazione

We move pieces of information across a multitude of latent spaces to concentrate them in segments that are useful for decision-making.

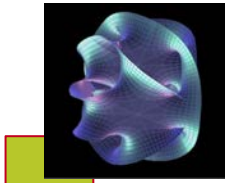
Training aims to discover these routes by exploring as many possibilities as possible through trial and error.

High and low dimensional spaces



We observe reality in large sparse spaces

We destroy the structure of information in even larger and sparser spaces



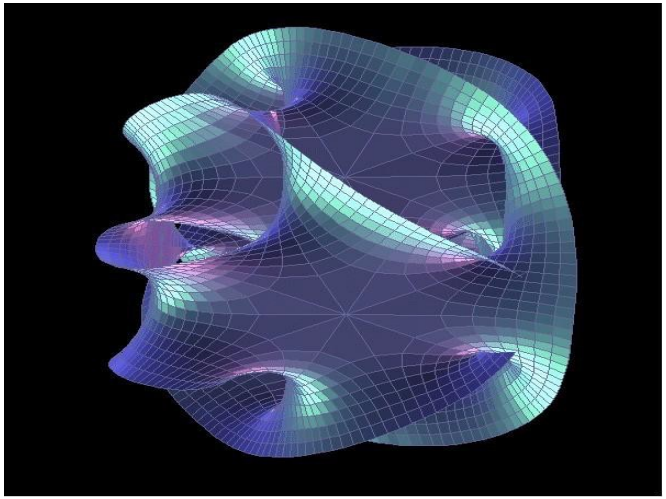
We discover the smaller non-linear denser generating space

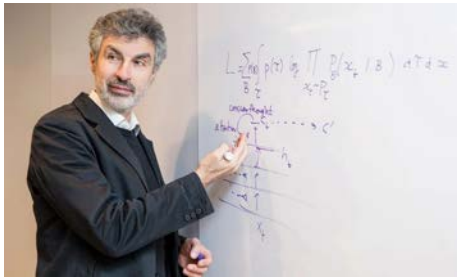
TWO MAIN CONSEQUENCES:

1. We need a lot of data and energy
2. Neural networks acts as (latent) memories

The manifold hypothesis

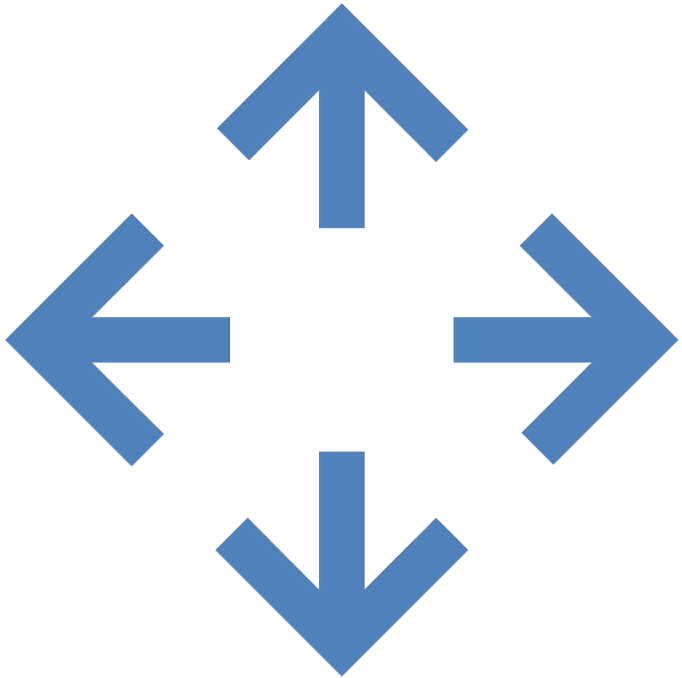
The manifold hypothesis suggests that high-dimensional data, often encountered in real-world applications, lies on or near a lower-dimensional manifold within the high-dimensional space, making it possible to effectively represent and analyze complex data with fewer dimensions.





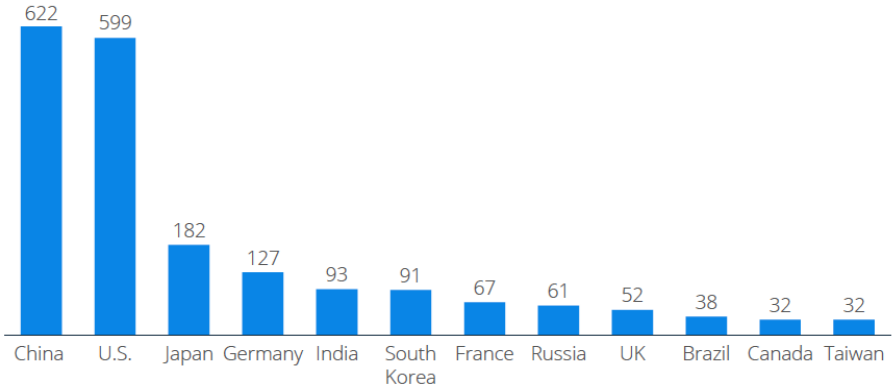
[Representation Learning: A Review and New Perspectives]

Quindi, che impatto avrà?

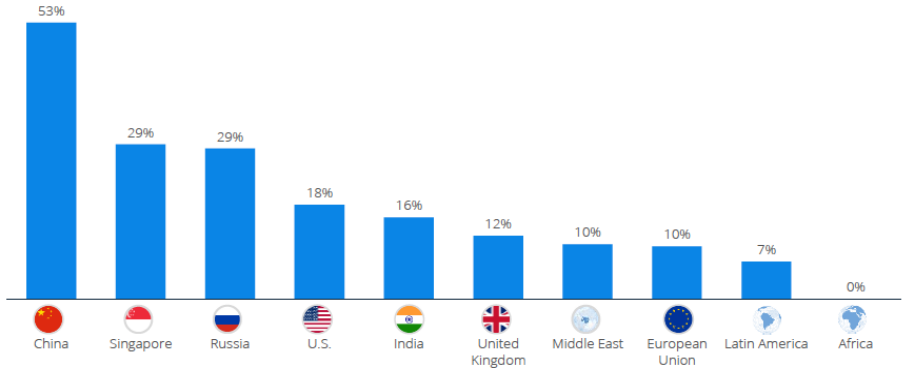


L'AI nel mondo non è tutta uguale

Leading countries by gross R&D expenditure in AI worldwide (2021, billion US$)



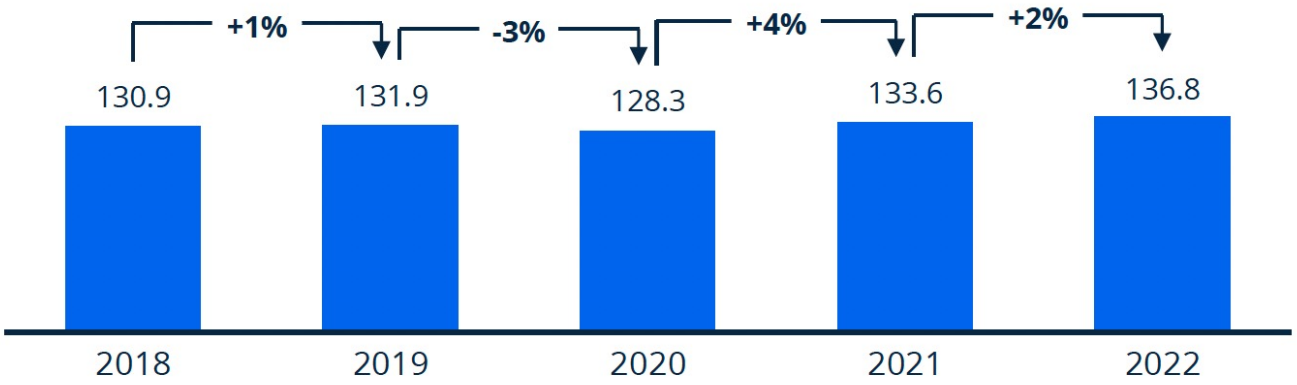
Expert estimation in the field of most impactful countries in AI innovation between 2022-2025[1]



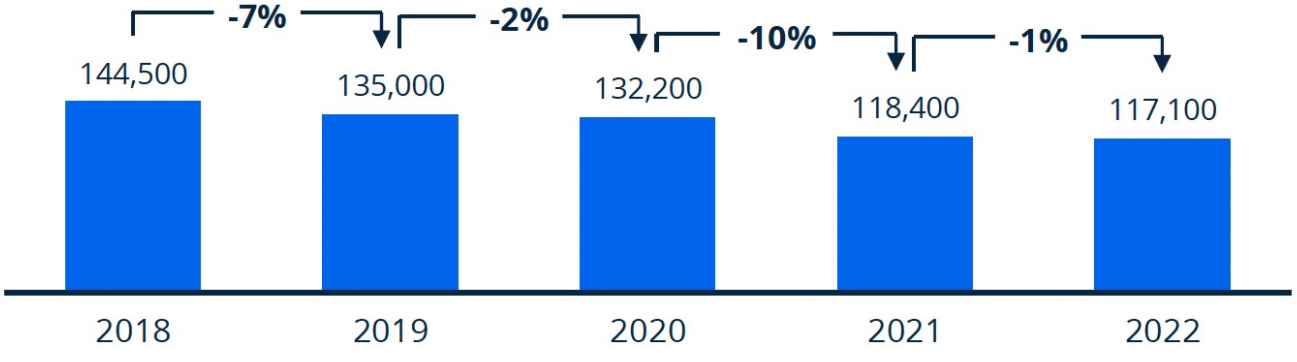
1: Survey conducted by Atlantic Council's GeoTech Center including more than 100 technology experts to record their expectations about the impact of COVID-19 on innovation in five key fields: the future of work, data and AI, trust and supply chains, space commercialization, and health and medicine
Sources: atlanticcouncil.org

Un esempio

Verizon global revenue[1] in billion US$



Verizon number of employees[2]



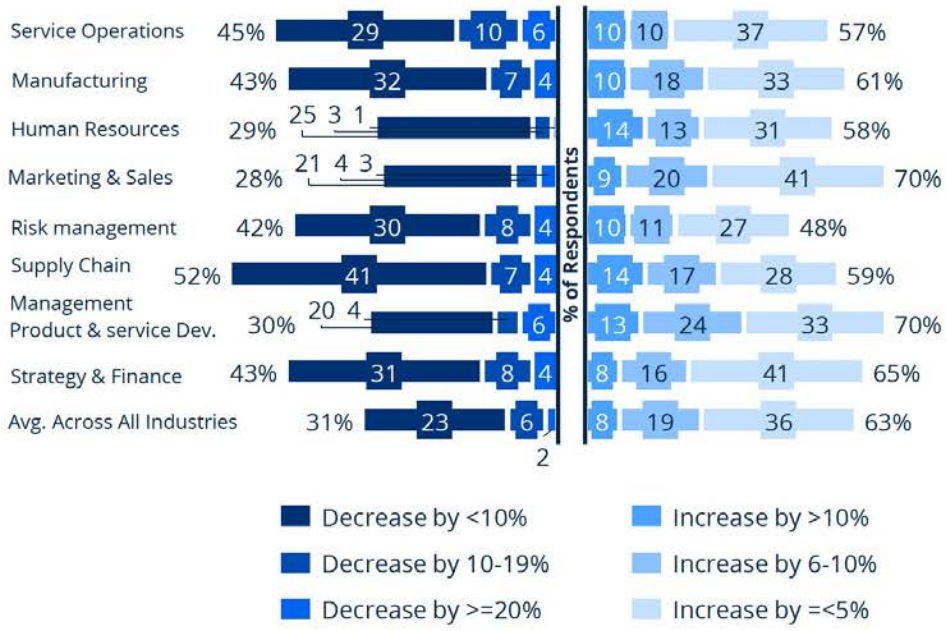
Su cosa impatterà l'AI?

Percent share of time spent on Microsoft 365[1]

Opinions of business leaders on adoption of AI in workplaces [2]

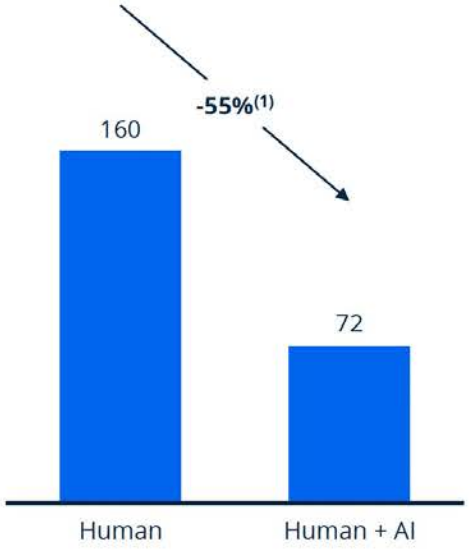
Su cosa impatterà l'AI?

Cost decrease and revenue increase from AI adoption by function in 2021



AI-Human productivity gains in selected professions

Time to complete coding tasks in minutes 2022 (1)



Time and cost to complete design tasks

Human
Cost: $150
Time: 5 hours

Generative AI
Cost: $0.08
Time: <1 minute

Daniel Newman identifies five perspectives on AI's societal impact:

- **Utopian Thought:** AI leads to unparalleled wealth and frees humans for meaningful work.
- **Dystopian Thought:** AI causes unemployment and economic issues, prompting suggestions like Universal Basic Income.
- **Tech-Optimistic Thought:** AI advancements will eventually result in significant productivity increases and economic benefits.
- **Realist Thought:** AI will be productive but may increase job polarization; informed decisions and research are necessary.
- **Lack of Productivity Thought:** AI won't significantly boost productivity due to various socio-economic challenges.

GRAZIE
PER LA VOSTRA
ATTENZIONE!