

Dall'AI Discriminativa all'AI Generativa

by Daniele Grotti

TABLE OF CONTENTS

01

Definizioni

02

Machine Learning

03

Ai Discriminativa

04

AI Generativa

05

Evoluzione degli
LLM

06

Q/A

Chi sono

- **Innovation Manager:** aiuto le aziende a navigare nel mondo digitale e a implementare nuove tecnologie legate all'intelligenza artificiale e all'automazione. Il mio lavoro consiste nel facilitare la collaborazione e l'integrazione delle idee innovative in tutti gli aspetti delle organizzazioni, dalla ricerca e sviluppo al marketing, dalla produzione alle vendite.
- **Data Scientist:** offro consulenza esterna per la creazione di pipeline per la pulizia dei dati, la trasformazione e la modellazione. La mia specialità è la consulenza strategica: utilizzo dati finanziari, di produzione ed open data per ottimizzare le strategie aziendali. Offro consulenza sulle opportunità di business, identificando aree di miglioramento e suggerendo soluzioni innovative.
- **Machine learning professor:** mi occupo di formazione post-laurea per il progetto Big Data Lab della regione ER dal 2018. Sono docente di corsi di programmazione Python e di algoritmi di ML e Deep Learning. Creo percorsi di formazione personalizzata per dirigenti e manager, aiutandoli a migliorare le loro competenze in questo campo in rapida evoluzione.



Definizione di dati

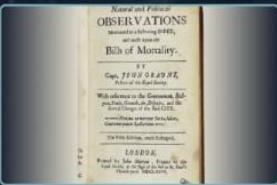
I dati sono rappresentazioni originarie, cioè non interpretate, di un fenomeno, evento, o fatto, effettuate attraverso simboli o combinazioni di simboli, o di qualsiasi altra forma espressiva legate a un qualsiasi supporto.

- Dati sono rappresentazioni di eventi o fatti
- Non interpretate (originarie)
- Attraverso simboli (o combinazioni di simboli)
- Contenute su supporti (forma espressiva)

HISTORY OF DATA



The Ishango bone holds the first evidence of data collection and storage.



John Graunt introduces the concept of data analysis in 1663.



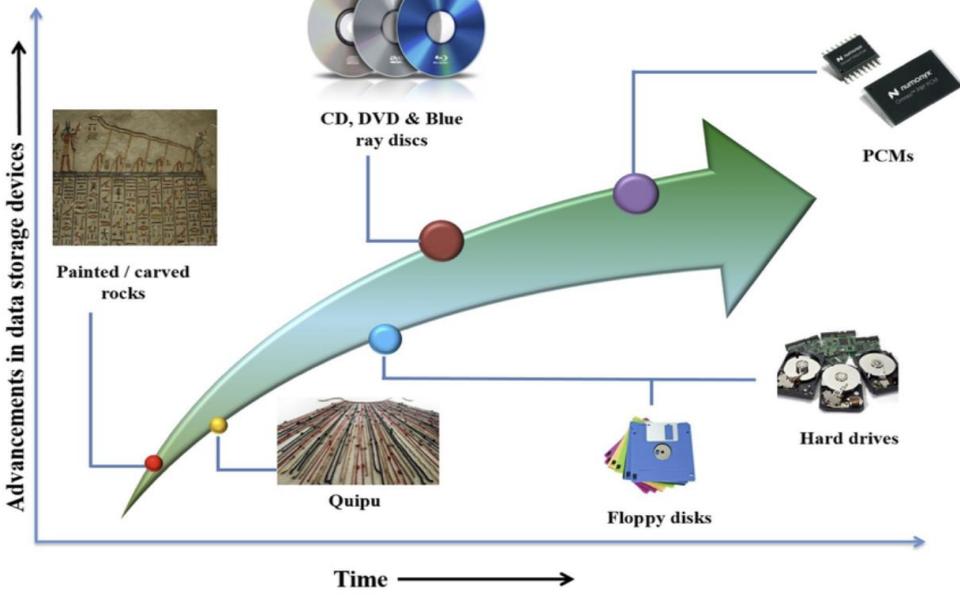
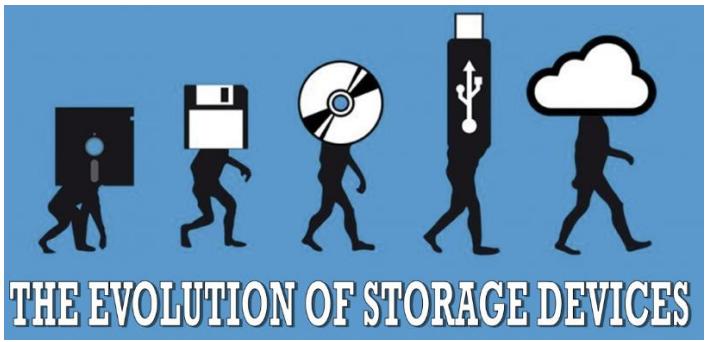
Herman Hollerith designs a machine that helped complete the US census in 1890.



Fritz Pfeleumer invents the magnetic tape which later inspired the invention of floppy disks and hard disk drives.



Sir Tim Berners Lee invents the World Wide Web.



Definizione di Informazione

L'informazione deriva da un dato, o più verosimilmente da un insieme di dati, che sono stati sottoposti a un processo di interpretazione che li ha resi significativi per il destinatario.

L'informazione è:

- insieme di dati
- interpretati
- comprensibili per il destinatario

DATI: DIAMANTI GREZZI



DATI

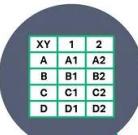


INFORMAZIONI

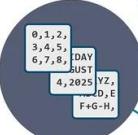


Structured Data vs Unstructured Data

Can be displayed
in rows, columns and
relational databases



Numbers, dates
and strings



Estimated 20% of
enterprise data (Gartner)



Requires less storage



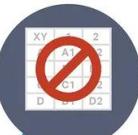
Easier to manage
and protect with
legacy solutions



vs

Structured Data vs Unstructured Data

Cannot be displayed
in rows, columns and
relational databases



Images, audio, video,
word processing files,
e-mails, spreadsheets



Estimated 80% of
enterprise data (Gartner)



Requires more storage



More difficult to
manage and protect
with legacy solutions



10%

Structured Data

Quantitative - in the form of numbers and values



90%

Unstructured Data

Qualitative - in the form of text files and varying formats

Structured Data

NPS/CSAT
CRM
Sales
Excel
Finance

Structured Data

Unstructured Data

Unstructured Data



Customer/Member Transactions



Online Communities



Chat



Notes & Text Fields

Surveys



Social Media



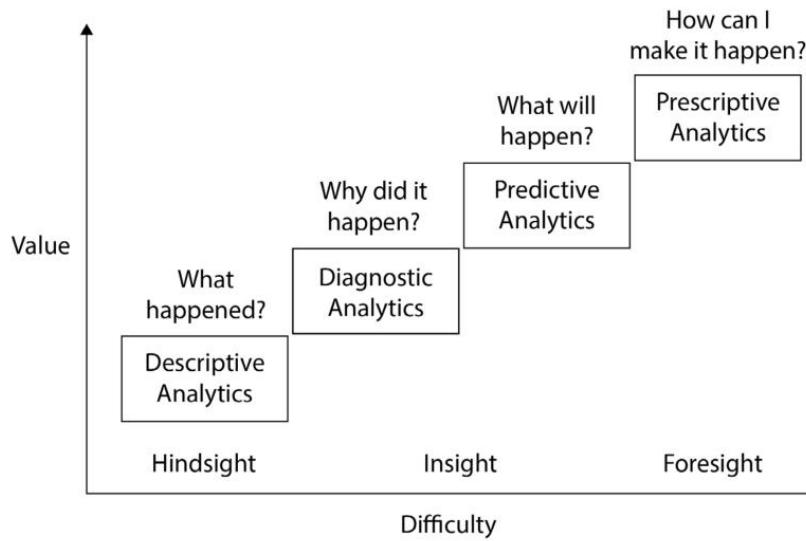
Voice Transcriptions

Call Center



Ratings & Reviews

ANALYTICS PATH TO VALUE (WITH DATA STORY)

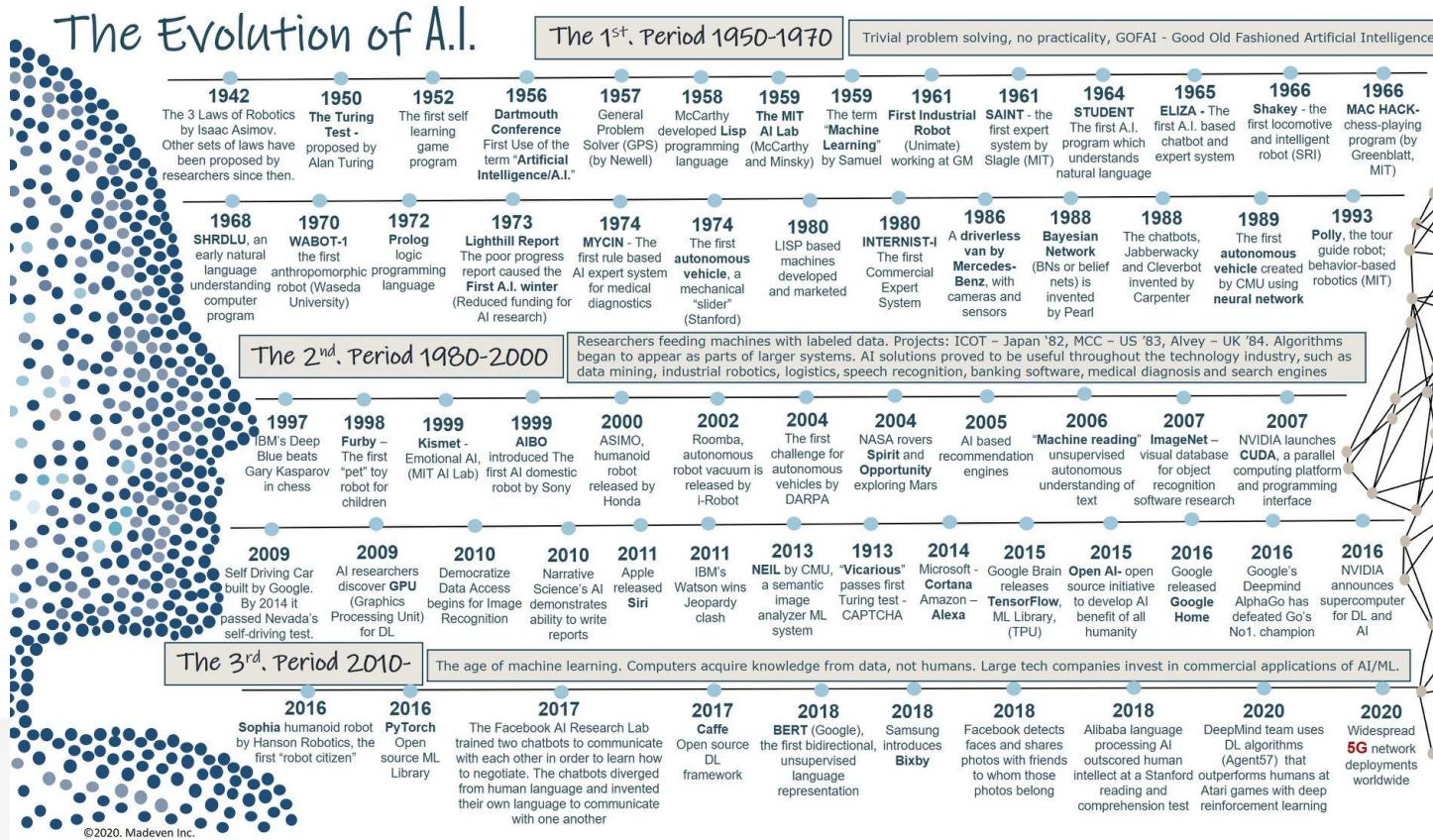


Cos'è AI?

AI: campo della scienza informatica che si occupa dello sviluppo di sistemi e algoritmi in grado di imitare l'intelligenza umana, consentendo alle macchine di apprendere, ragionare, risolvere problemi e prendere decisioni in modo autonomo

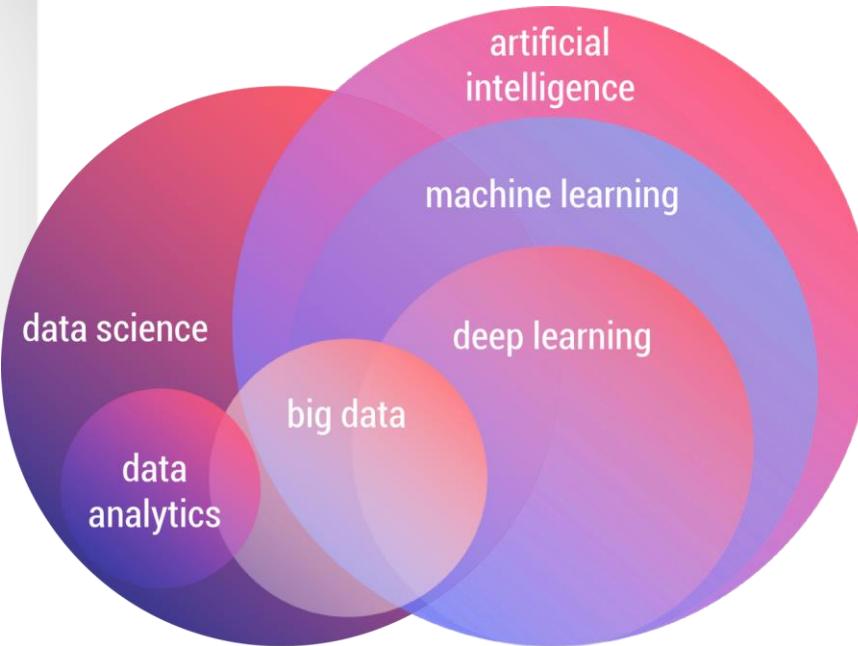
L'obiettivo principale dell'AI è quello di creare macchine in grado di imitare e superare le capacità cognitive umane in diversi ambiti

Da quanto tempo esiste AI?



Cos'è il Machine Learning?





"E' il settore dell'intelligenza artificiale che studia come dare ai computer l'abilità di imparare senza essere esplicitamente programmati" - 1959

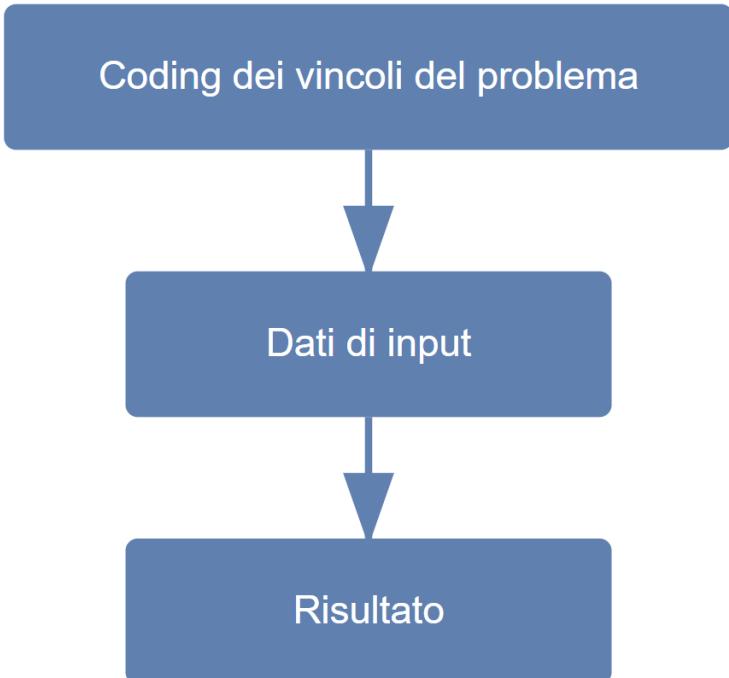


ARTHUR SAMUEL

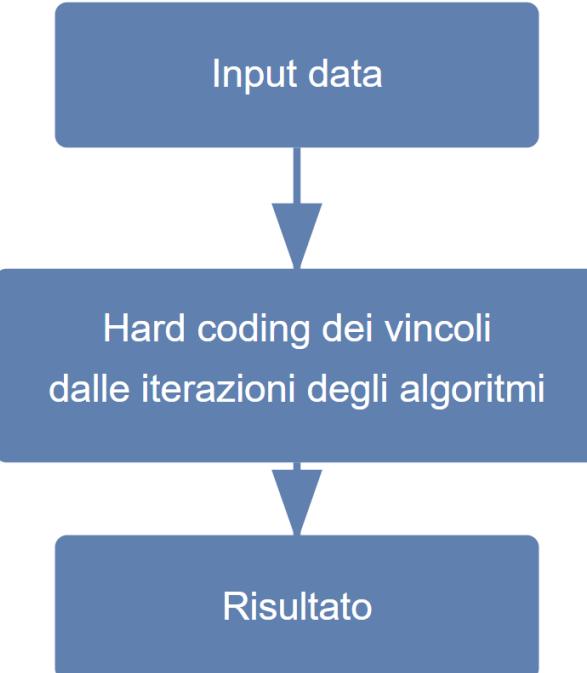
Pioniere dell'AI
Inventore del termine "Machine learning"

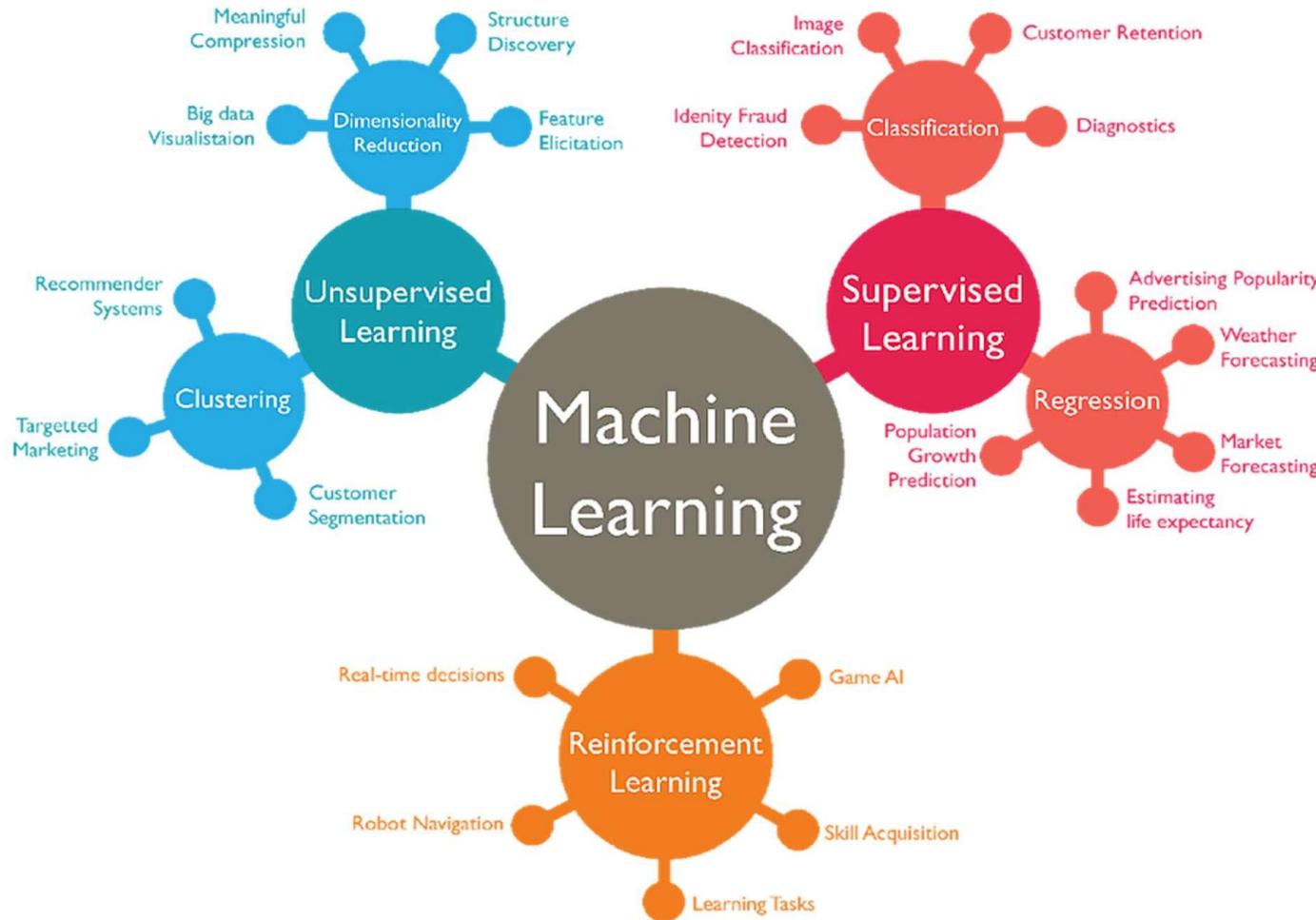


Programmazione classica



Machine learning

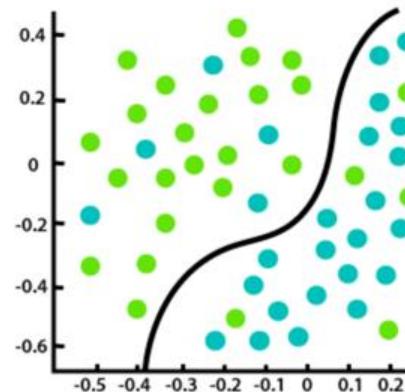




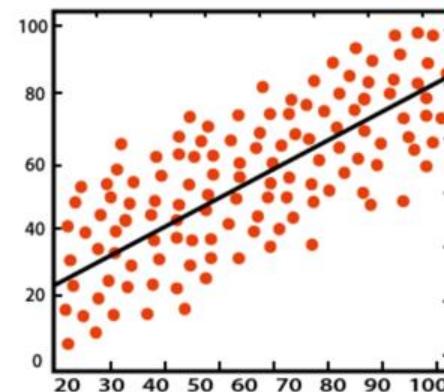
SUPERVISIONATO



Classification



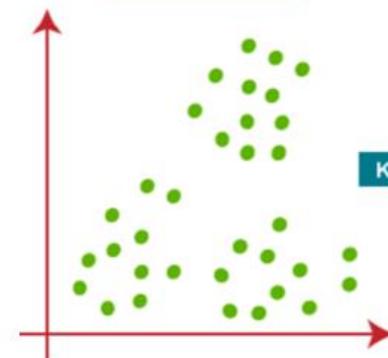
Regression



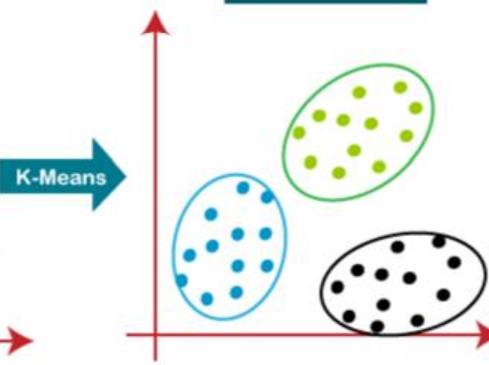
NON SUPERVISIONATO



Before K-Means

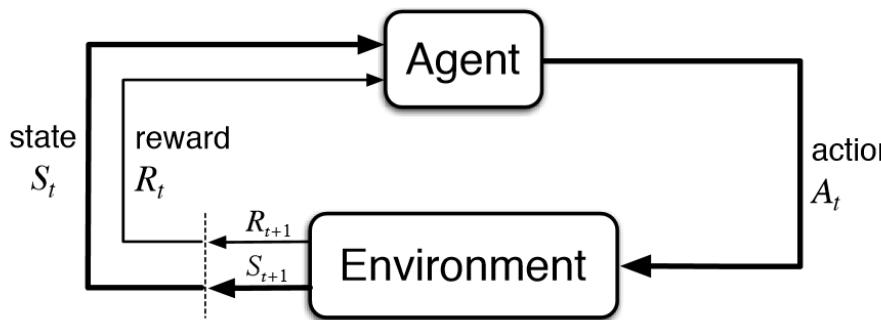
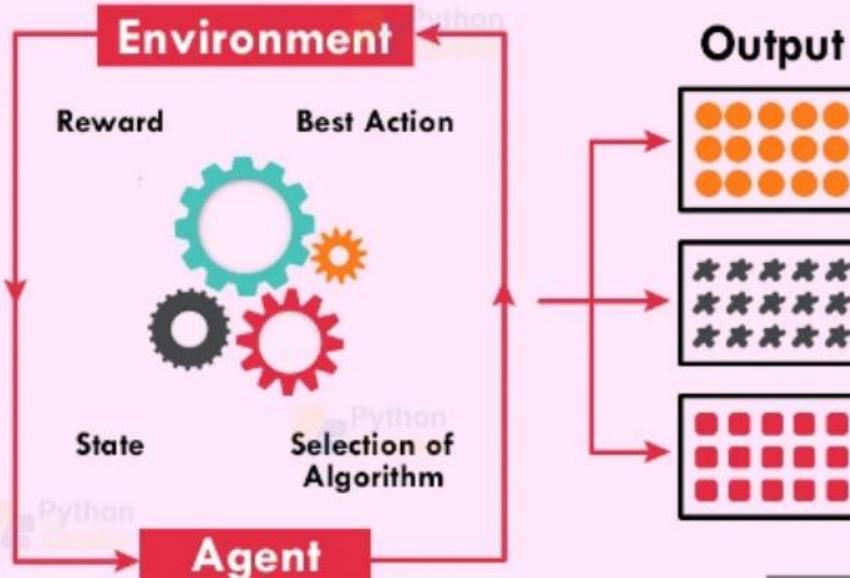
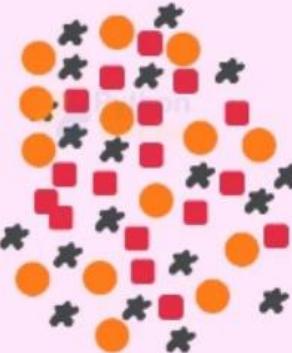


After K-Means



Reinforcement Learning in ML

Input Raw Data



DISCRIMINATIVE

VS.

GENERATIVE

Discriminative Model

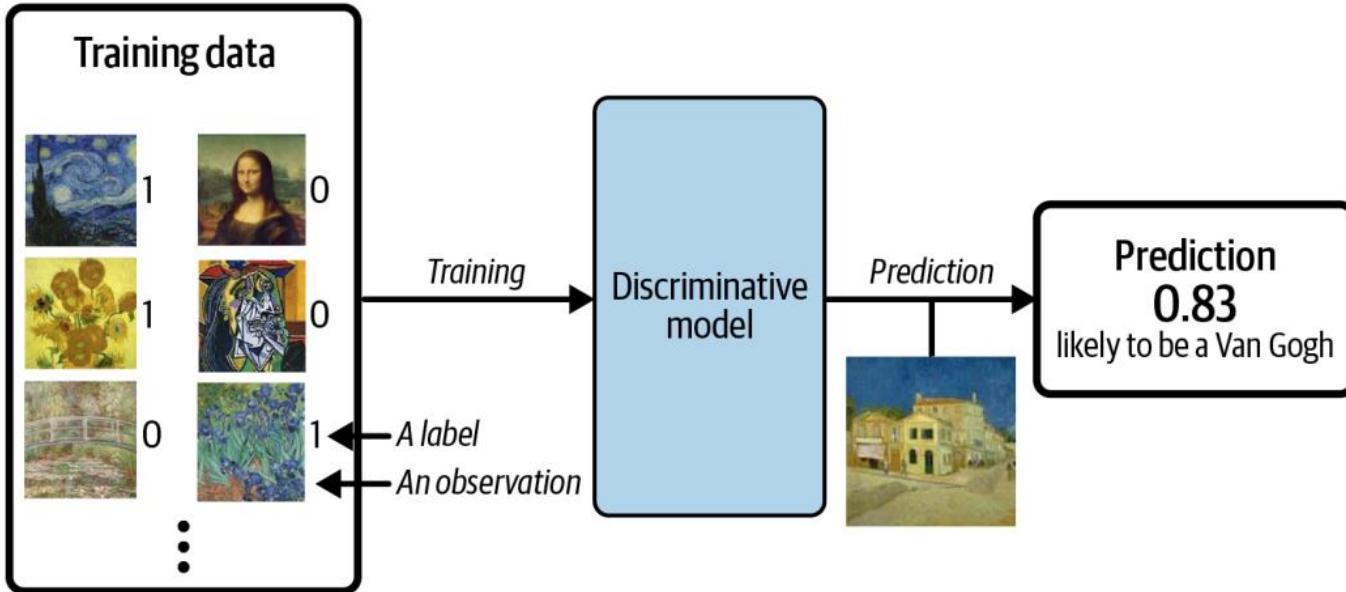


Figure 1-2. A discriminative model trained to predict if a given image is painted by Van Gogh

What Is Generative Modeling?

Generative modeling can be broadly defined as follows:

Generative modeling is a branch of machine learning that involves training a model to produce new data that is similar to a given dataset.

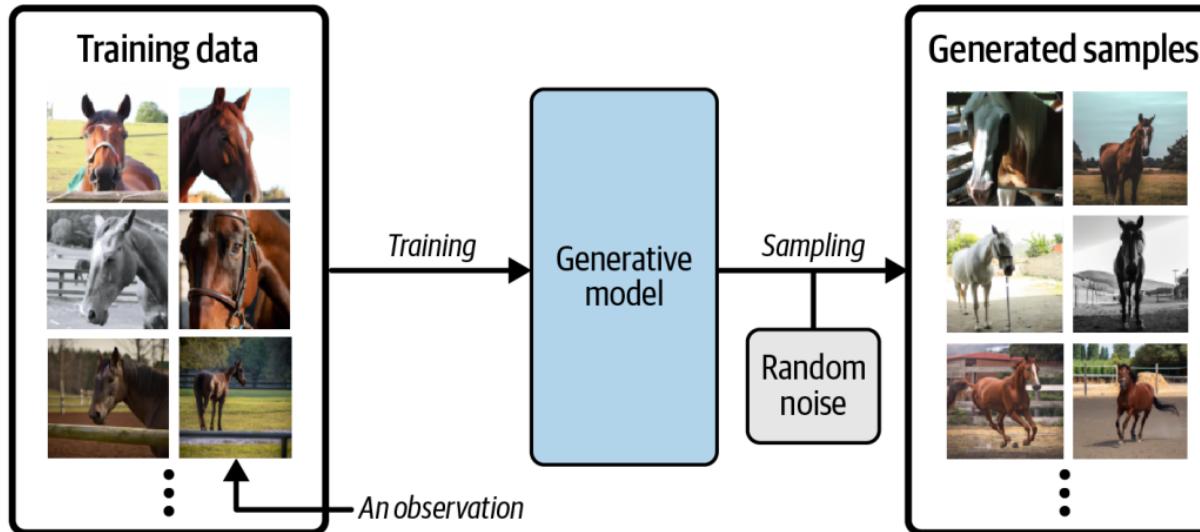
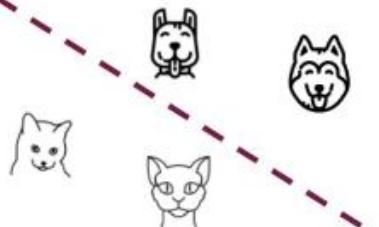


Figure 1-1. A generative model trained to generate realistic photos of horses

Generative Models vs. Discriminative Models

Discriminative models



Features Class

$$X \rightarrow Y$$

$$P(Y|X)$$

Generative models

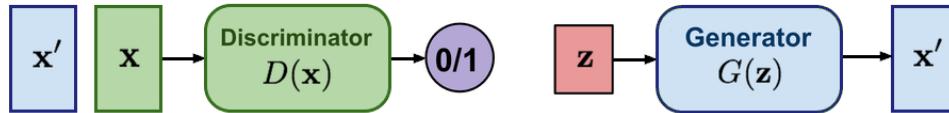


Noise Class Features

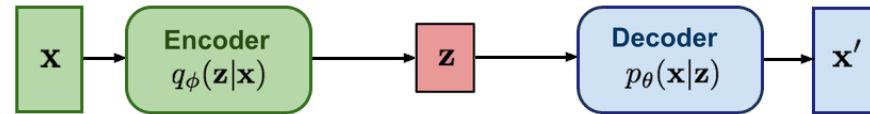
$$\xi, Y \rightarrow X$$

$$P(X|Y)$$

GAN: Adversarial training

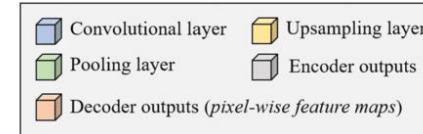
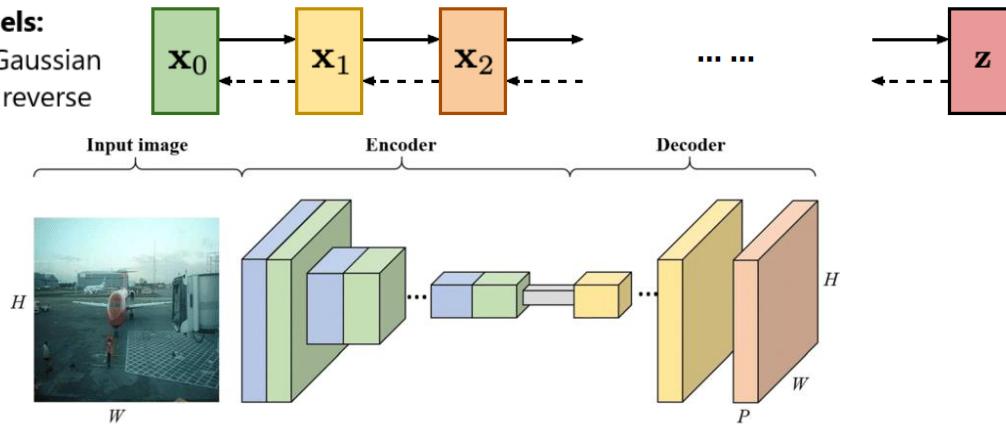


VAE: maximize variational lower bound



Diffusion models:

Gradually add Gaussian noise and then reverse



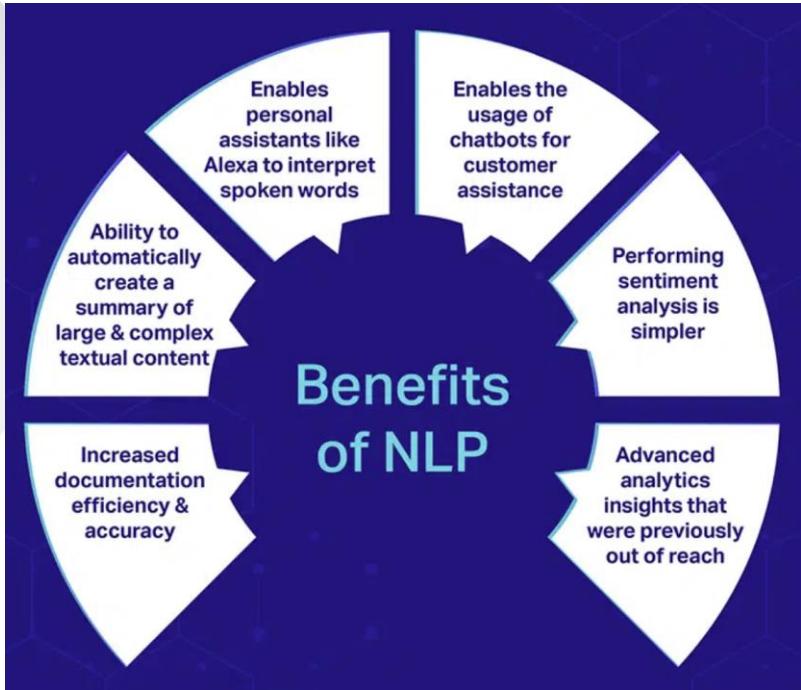


Prompt: Epic anime artwork of a wizard atop a mountain at night casting a cosmic spell into the dark sky that says "Stable Diffusion 3" made out of colorful energy



NLP Market Size & Growth

NLP market is set to reach **\$156.80 billion** by 2030, growing at a CAGR of **27.55%**. Over **85%** of large organizations plan to adopt NLP by 2025, driven by various factors.



Challenges



Misspellings



Language Differences



Innate Biases



Words with
Multiple Meanings



Uncertainty and
False Positives



Training
Data

Application of Text Classification

Sentiment Analysis

Text sentiment, categorizing content as positive, negative, or neutral.

Spam Detection

Identify and filter out spam emails or messages by analyzing their content and characteristics, enhancing communication security.

Topic Labeling

Automatically assigning topics or categories to documents, making content organization and retrieval more efficient.

Language Identification

Detect the language in which a text is written, which is useful for multilingual content processing and translation.

News Categorization

News articles to be categorized into sections like politics, technology, sports, etc., improving content organization for readers.

Product Classification

E-commerce platforms use NLP for product categorization, ensuring items are correctly labeled and presented to customers.

Customer Feedback Analysis

Customer reviews and feedback to extract insights, understand customer satisfaction and areas of improvement.

Medical Document Classification

Categorizing medical records, research papers, and patient notes, assisting in efficient data retrieval for healthcare professionals.

Legal Document Categorization

Law firms use NLP to classify legal documents, making managing and retrieving information from large databases easier.

Social Media Monitoring

NLP tracks and classifies social media posts, tweets, and comments, allowing brands to monitor their online presence and engage with users.

Fraud Detection

Classify financial texts to detect fraudulent activities and identify potential risks.

Resume Screening

Resume screening by categorizing job applications based on skills, experience, and qualifications.

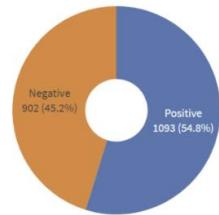
Content Recommendation

Recommending relevant articles, blogs, or products to users based on their interests.

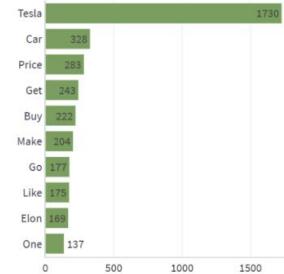


All Positive 😊 Negative 😥

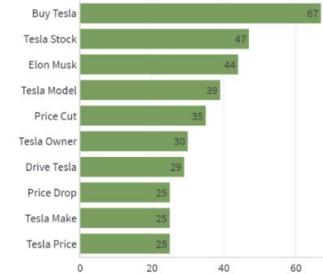
Sentiment Distribution



Top 10 Occuring Words



Top 10 Occuring Bigrams

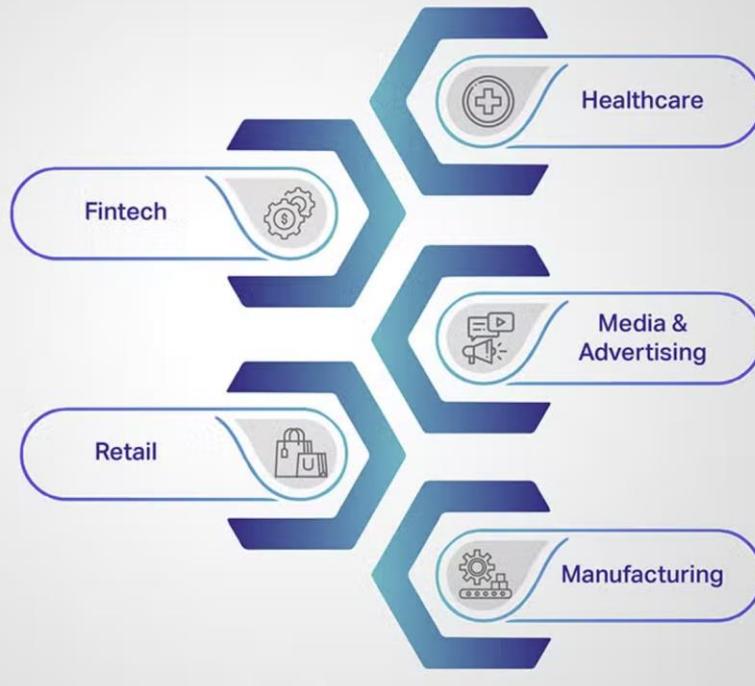


	Sentiment	Tweet
0	Negative	@YatinKheti @scidood Again it sounds like some limited release stuff and not re
1	Negative	@3RumsDeep @Tesla @elonmusk @montana_skeptic any bets on if this has ar
2	Negative	@ValueAnalyst1 Tesla's demo was obviously hacked together. You can't be that
3	Positive	@rmcadory @farzyness @GM @Toyota @Ford @Tesla Tesla doubling every 12 t
4	Negative	Well good news from yesterday is the amount I would get for my truck is higher
5	Negative	@alphadoge777 @elonmusk @Tesla @mayemusk Exactly.
6	Positive	I don't like asking for many favors. But this is a time where I am going to need y
7	Negative	@farzyness Redo: you need to compare global to global. Tesla sells 1.3m globa
8	Positive	@stevenward @Tesla Sweet feelings:) Such a lovely pic 🌸

Wordcloud



Industries Leveraging NLP



Healthcare

NLP offers rewarding benefits to the healthcare industry such as:

- the extraction insights from medical records and analysis of unstructured data
- Improve and personalize clinical decision support systems
- Optimize responses from chatbots for seamless patient care experiences
- Monitor, predict, and mitigate adverse drug reactions and implement pharmacovigilance strategies and more

Fintech

The implications of NLP in fintech is completely different, offering benefits like:

- Seamless document processing and onboarding
- Optimize risk management and fraud detection
- Assessment of creditworthiness of individuals for financing
- Personalization of financial products in terms of tenures and premiums and more

Embeddings



Embedding

anatine amigos

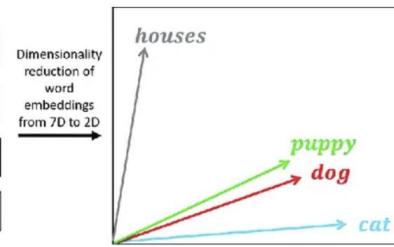
Text

Embedding
model

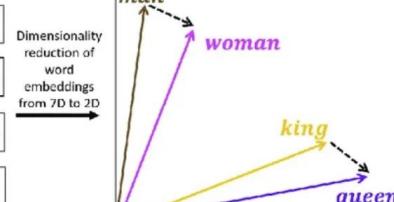
-0.027 -0.001 -0.020 ... -0.023

Text as vector

	d1	d2	d3	d4	d5	d6	d7
dog →	0.6	0.9	0.1	0.4	-0.7	-0.3	-0.2
puppy →	0.5	0.8	-0.1	0.2	-0.6	-0.5	-0.1
cat →	0.7	-0.1	0.4	0.3	-0.4	-0.1	-0.3
houses →	-0.8	-0.4	-0.5	0.1	-0.9	0.3	0.8

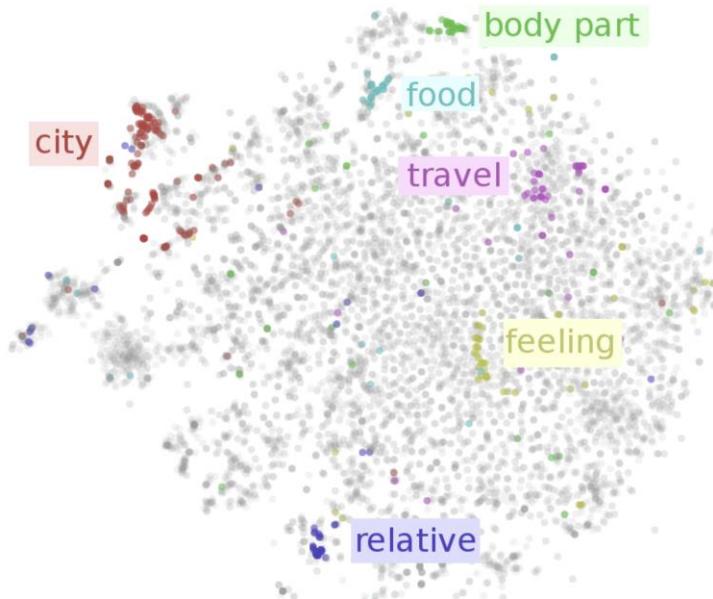


man →	0.6	-0.2	0.8	0.9	-0.1	-0.9	-0.7
woman →	0.7	0.3	0.9	-0.7	0.1	-0.5	-0.4
king →	0.5	-0.4	0.7	0.8	0.9	-0.7	-0.6
queen →	0.8	-0.1	0.8	-0.9	0.8	-0.5	-0.9



Word Word embedding Dimensionality reduction

Visualization of word embeddings in 2D



<https://projector.tensorflow.org/>

Transformers

6 Dec 2017

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

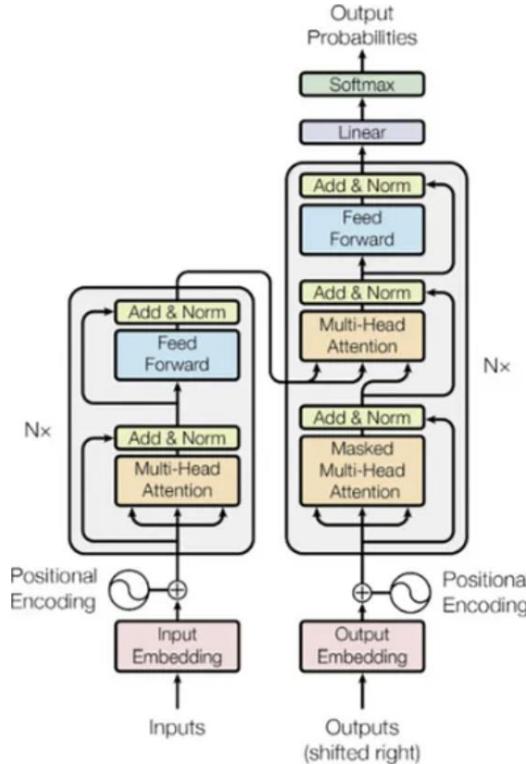
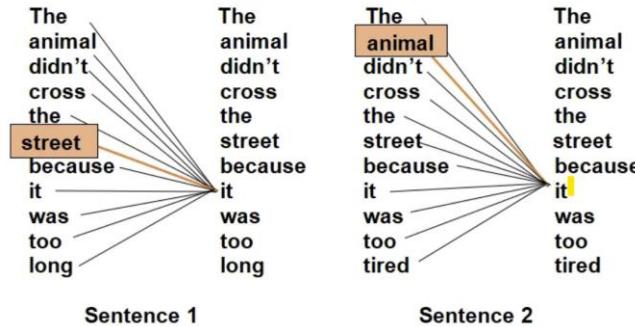
Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

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

Lukasz Kaiser*
Google Brain
lukaszkaiser@google.com

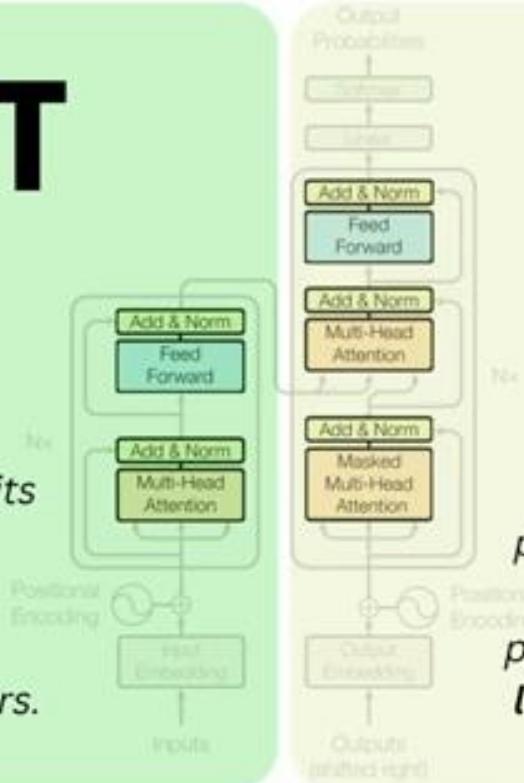
Illia Polosukhin* ‡
illia.polosukhin@gmail.com



BERT

Google

*use transfer learning to **continue** learning from its existing data when adding user-specific tasks and layers.*



GPT

OpenAI

*decodes from its massive pre-learned embeddings to present output that matches user prompts. It does **not** learn anything new.*

LLM

LARGE LANGUAGE MODEL

Text Training Data

Web Data

Books

Images converted to text

Structural Data



Training

Large Language Model

Adaption

Adaptations

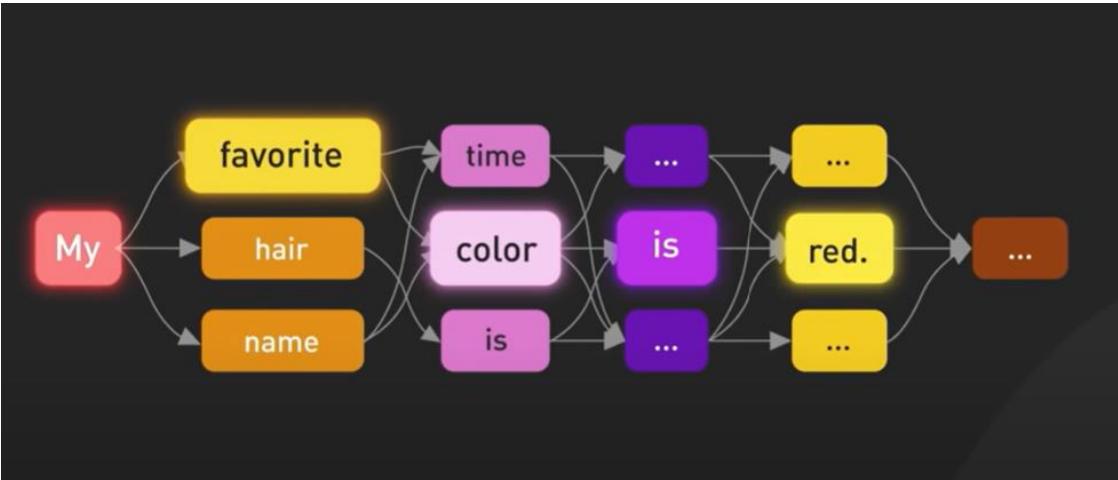
Question Answering

Sentient Analysis

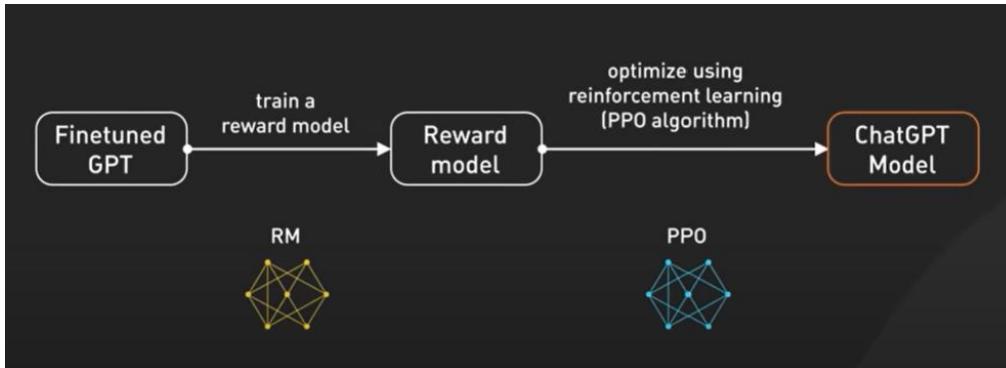
Information Summarization

Planning

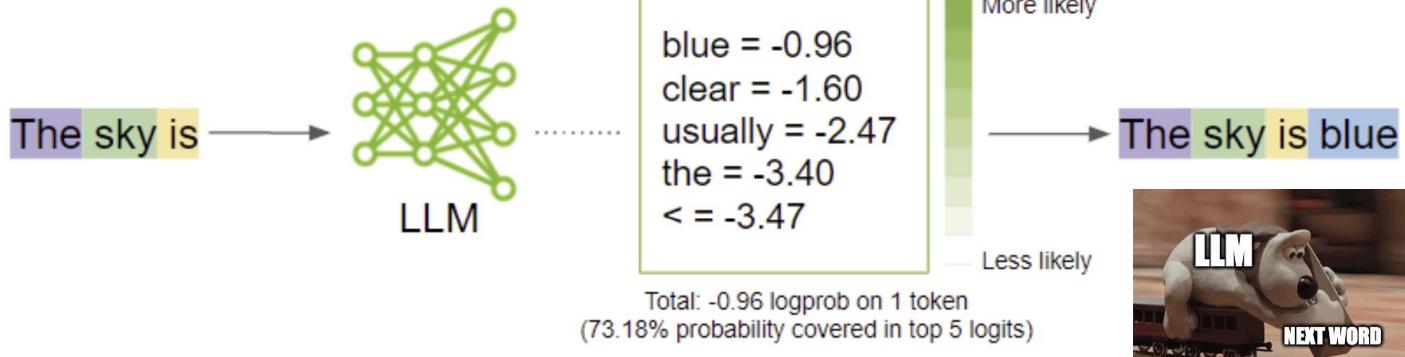
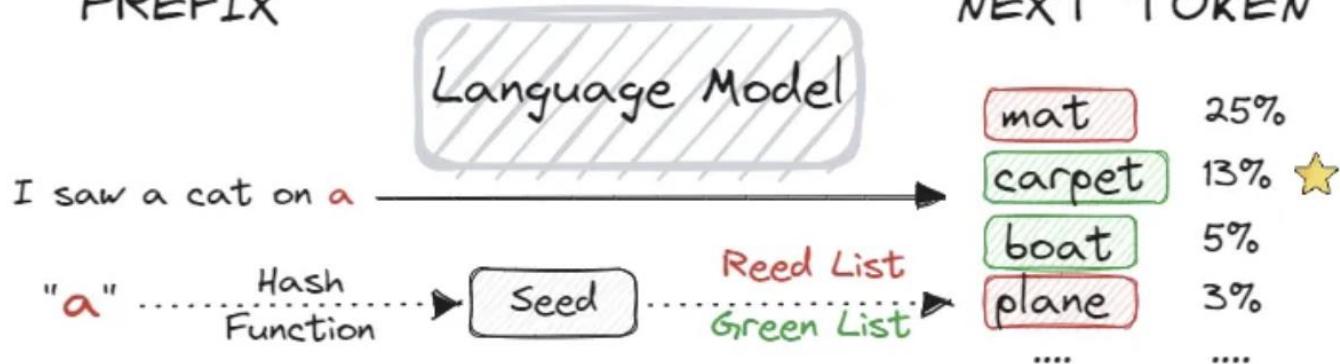
Language Translation

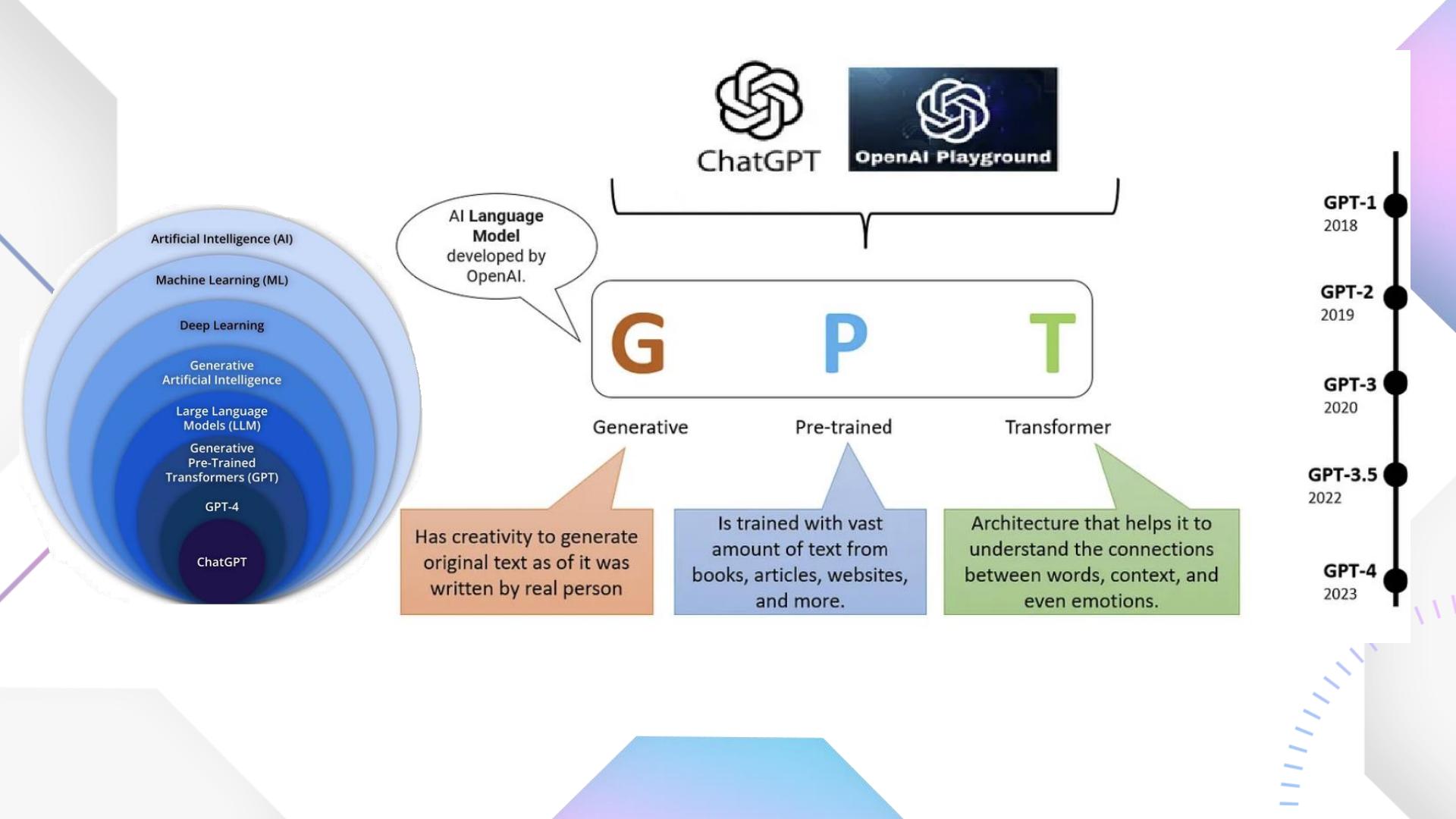


Reinforcement Learning with Human Feedback



PREFIX





Generative AI

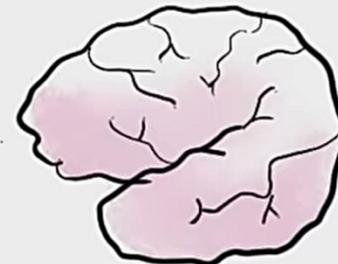
Computers have gotten smarter

Execute Instructions

```
function withdraw(amount) {  
    if (amount > balance) {  
        fail("Hey you ain't got the cash!")  
    } else {  
        balance = balance - amount  
    }  
}
```



Learn + Think





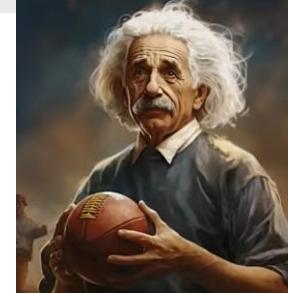
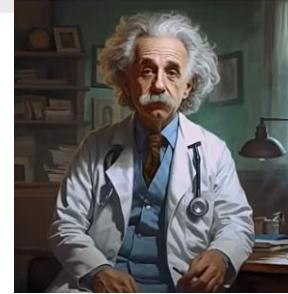
 **You**
What are you? Explain in 1 sentence to a 10 year old.

 **ChatGPT**
I'm like a super smart robot that can answer lots of questions and help with all sorts of things, kind of like a really helpful genie in a computer!

Einstein in your basement



A cartoon illustration of Albert Einstein sitting cross-legged on the floor, looking up at a simple light fixture. He has his signature wild white hair and a thoughtful expression. Above him is a speech bubble containing the text "Einstein in your basement".



Limiti umani:

- Fare errori
- Giungere a conclusioni affrettate
- Può faintendere
- Inventarsi soluzioni che non esistono
(allucinazioni)

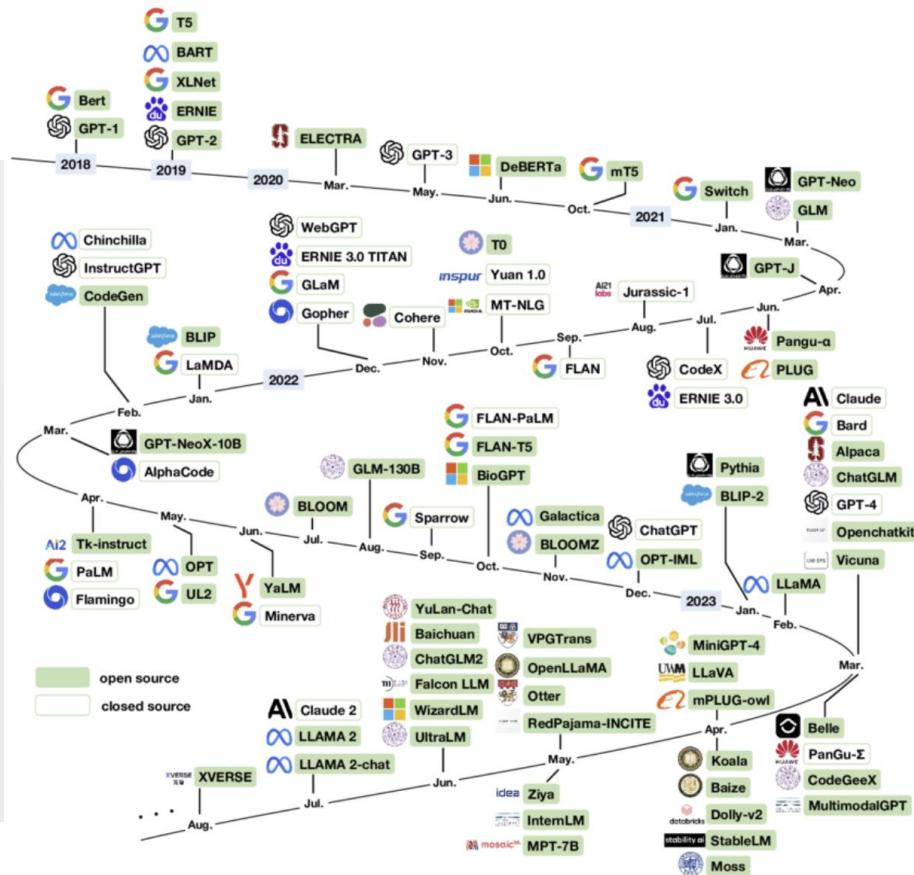
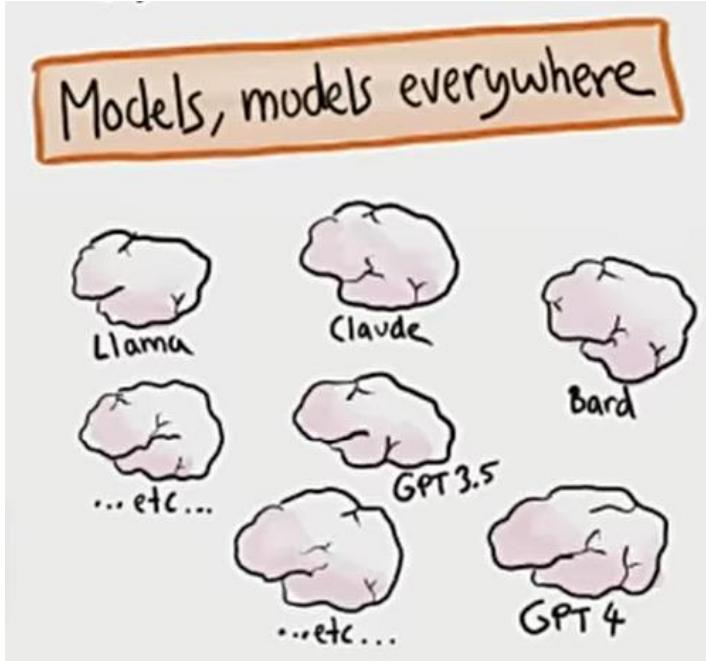
Biggest limitation
is you

Imagination

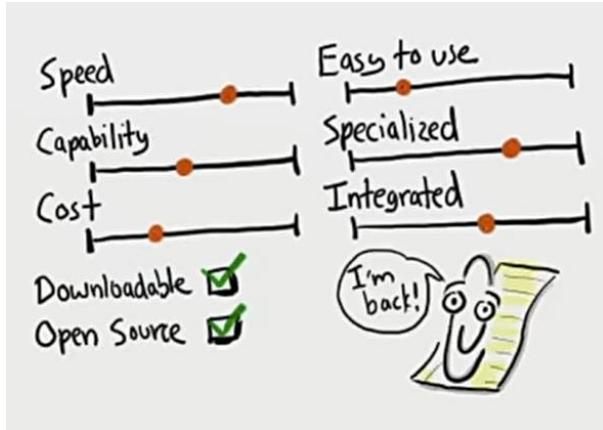
What can
I do?

Prompt
Engineering
skills

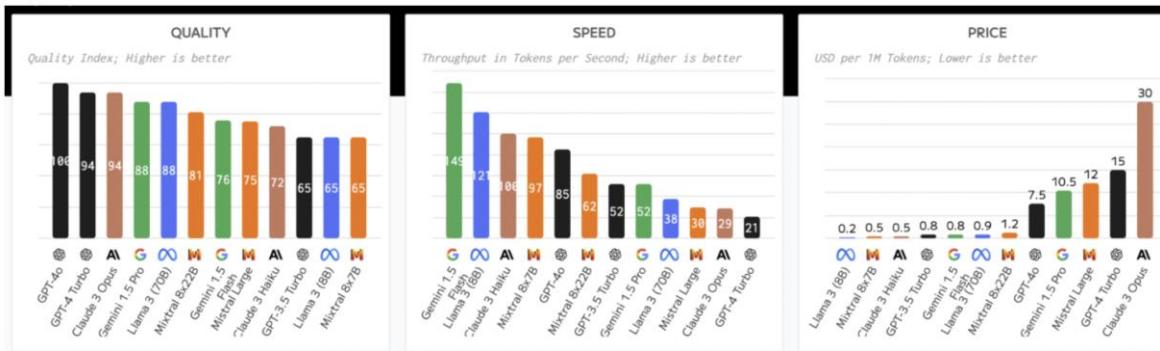
How do
(I do it?)



Open-Source Vs. Close-Source LLMs



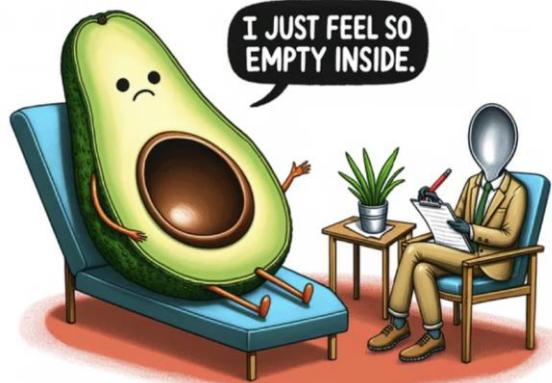
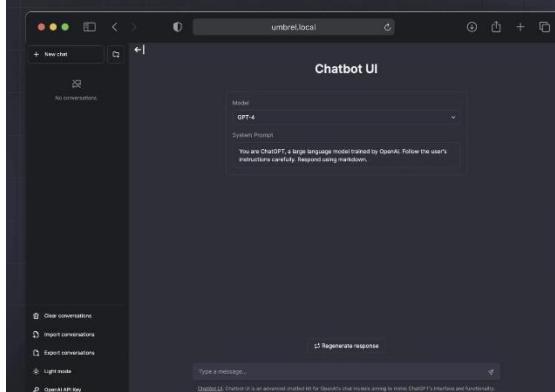
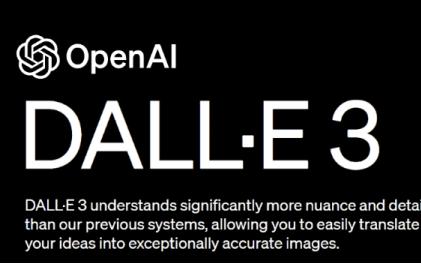
Features	Open-Source LLMs	Close-Source LLMs
Cost	Lower initial cost but possibly higher maintenance.	Higher initial cost but lower maintenance.
Innovation	Collaborative and fast innovation from a global community.	Slower innovation directed by a single entity.
Support	Reliant on community-based support, which can be extensive but inconsistent.	Consistent and quality support offered through dedicated channels.
Customization	Highly customizable with access to source code.	Customization possible within the platform's limits.



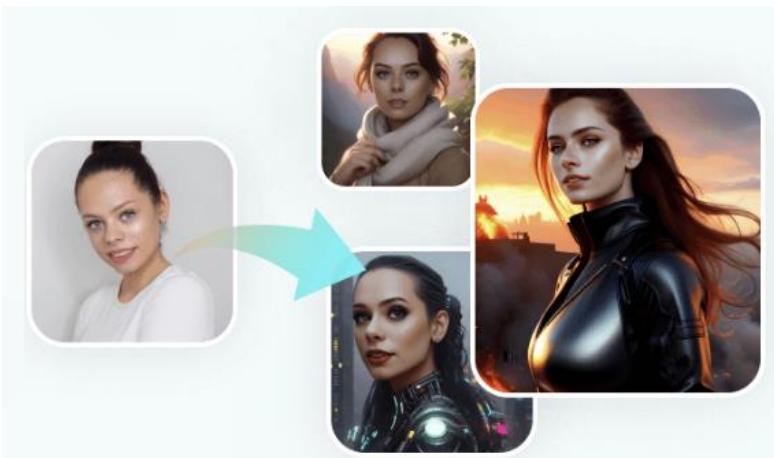
Model Types

Text to Text

Text to Image



Model Types



What are we looking at here?

✓ Analyzing the image: Privacy blur hides faces from Bing chat

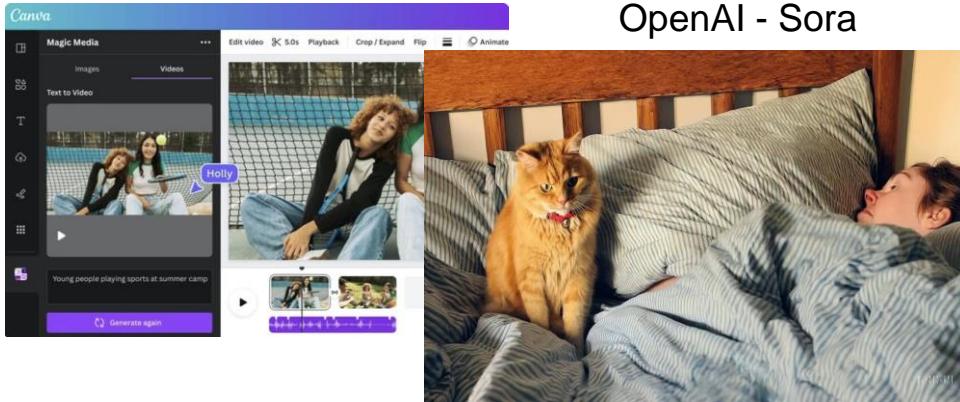
The image you're seeing is quite intriguing. It features a person sitting cross-legged on a grassy field, holding a large, oval-shaped mirror with a gold frame. The person's face is blurred out, maintaining their privacy. The mirror reflects the surrounding environment, adding an interesting perspective to the scene. The backdrop of the image is a grassy field with trees in the distance, contributing to the overall serene and tranquil atmosphere. The image has a dark and moody tone, which adds a layer of depth and emotion to it. It's a beautiful blend of nature and human presence, don't you think?

Model Types

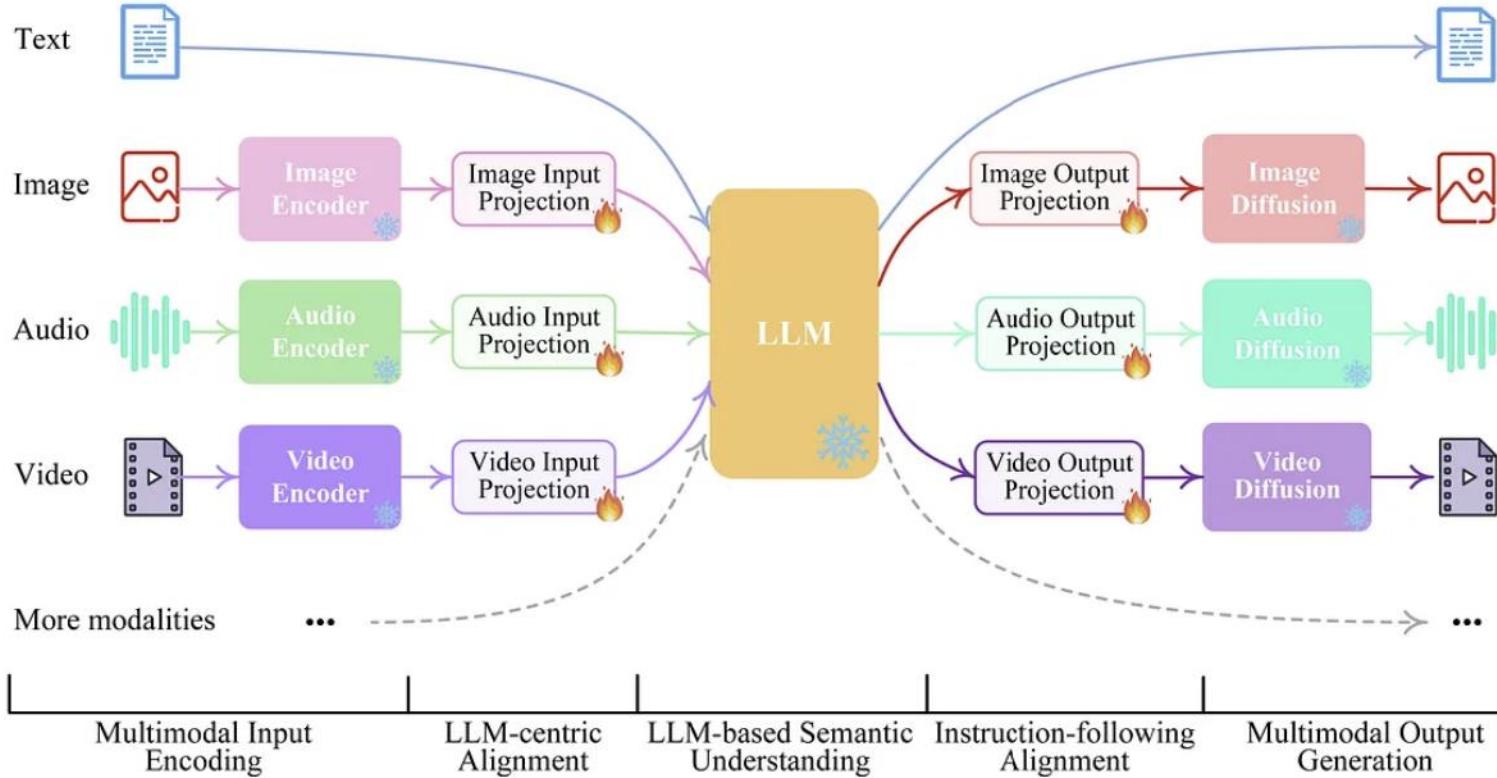
Text to Video

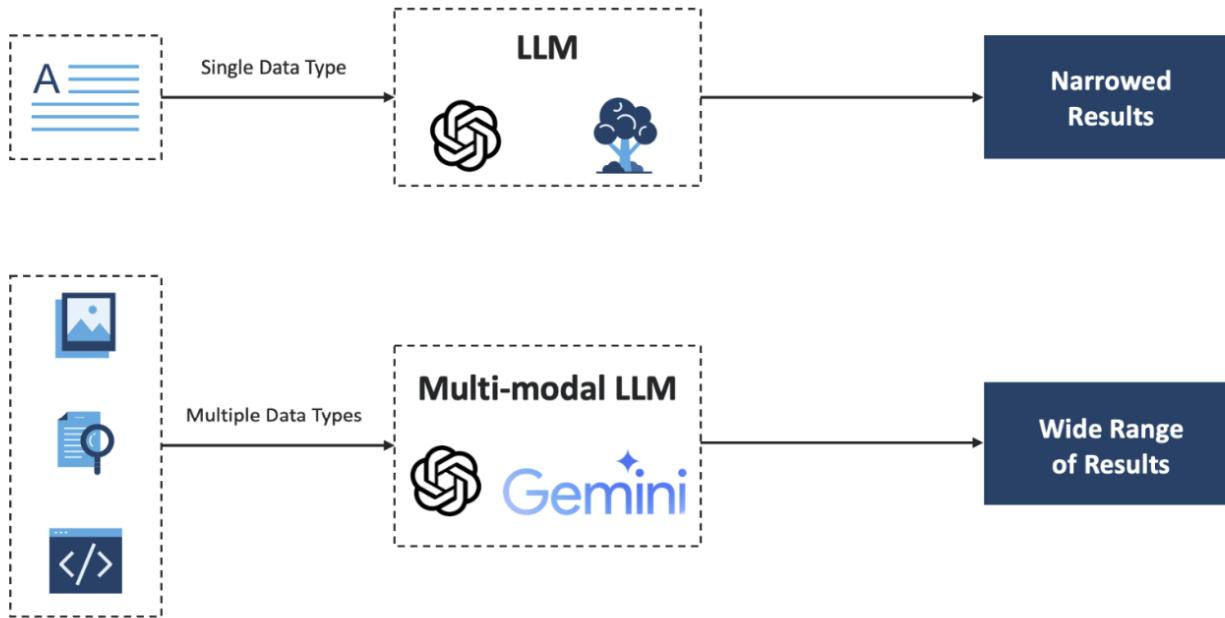
Speech to Text

Text to Speech



Multi- Modality





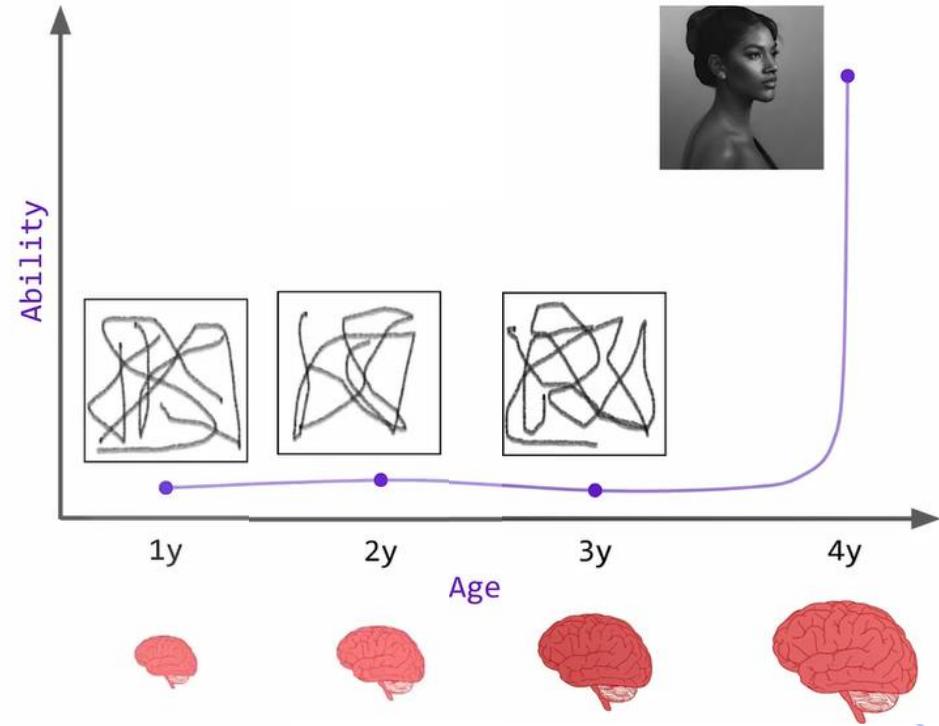
Comprendere contestuale

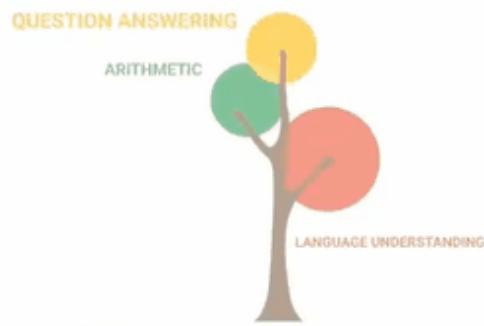
- Analisi simultanea di testo e immagini
- Interpretazione più accurata del contesto completo

Versatilità applicativa

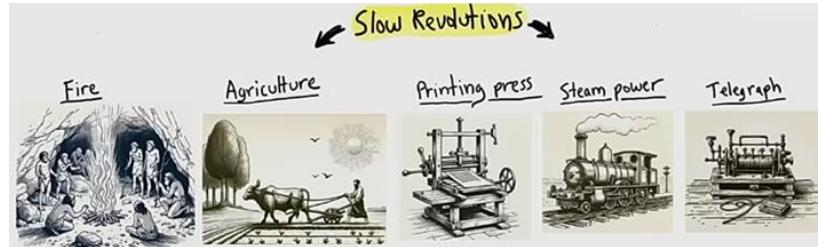
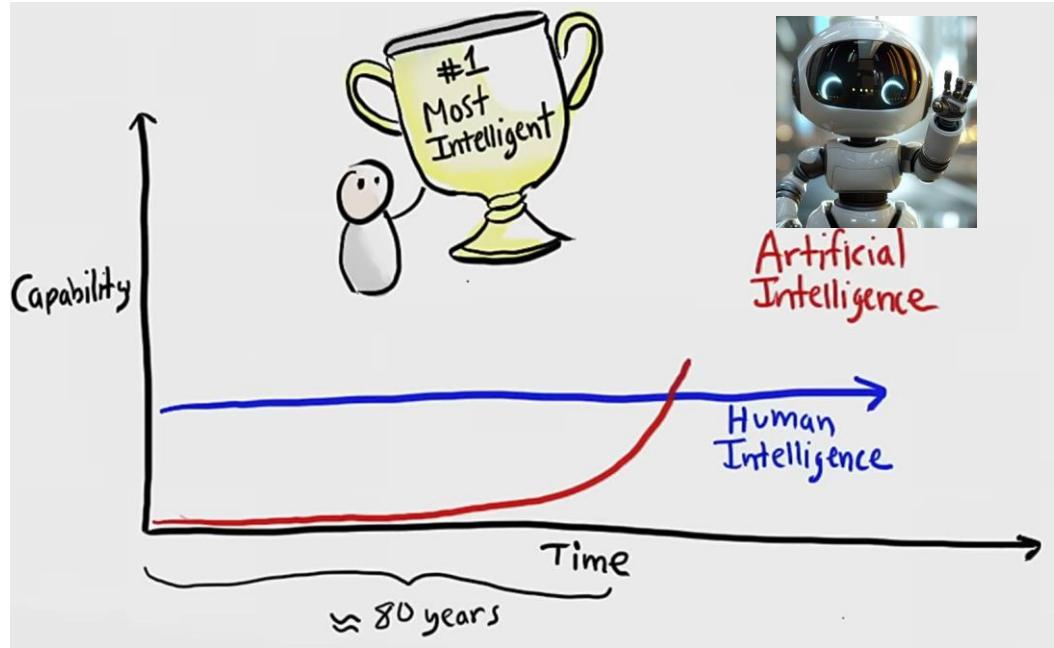
- Assistenza visiva in vari campi (medicina, architettura)
- Supporto decisionale basato su dati multimodali

The Emergent Abilities of LLMs

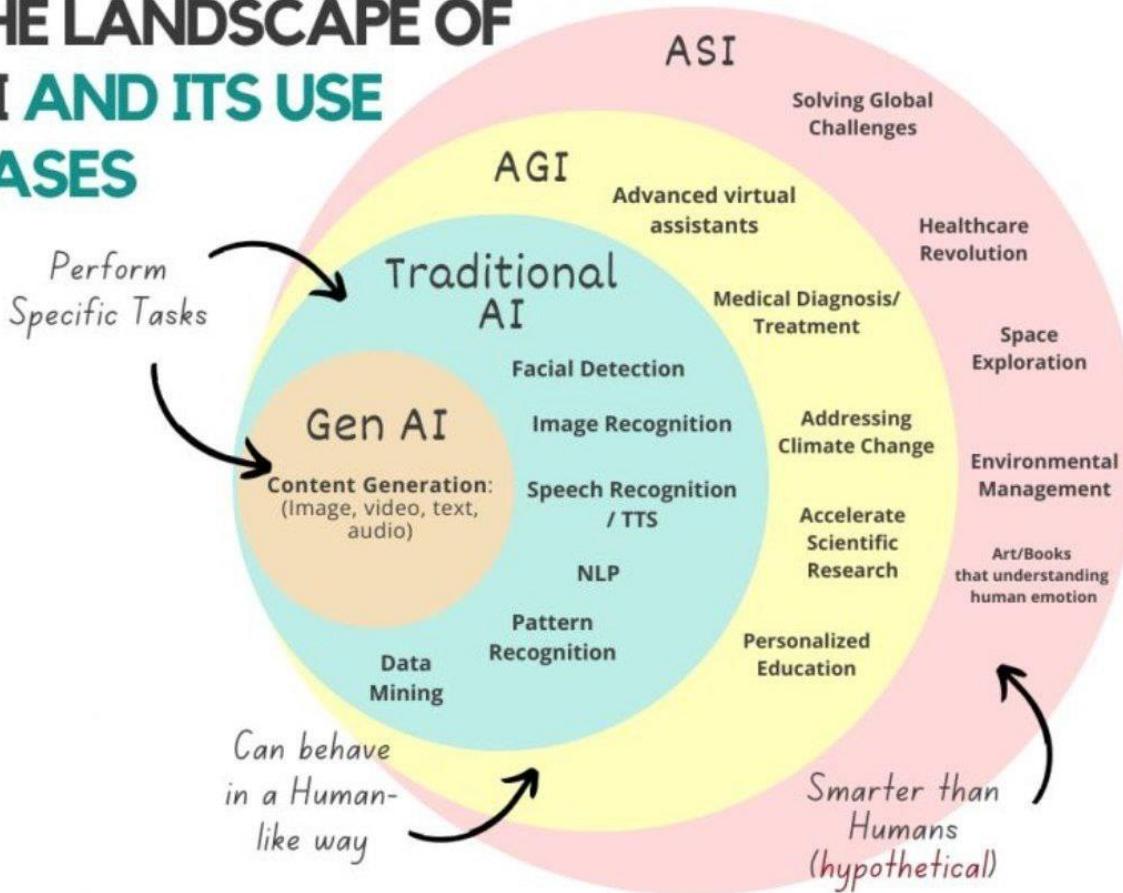




8 billion parameters

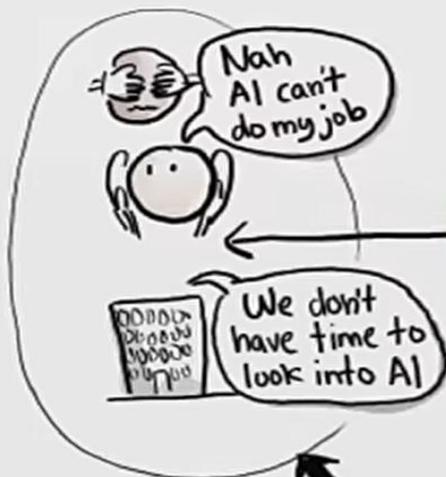


THE LANDSCAPE OF AI AND ITS USE CASES



Mindset

Denial



Positive



Panic



AI might not take your job,
but people/companies using AI will

The role of Humans

Is human role X still needed?

doctor, developer, lawyer, ceo, teacher, etc...

Decide what to ask
and how

Provide context

Legal compliance

Data security

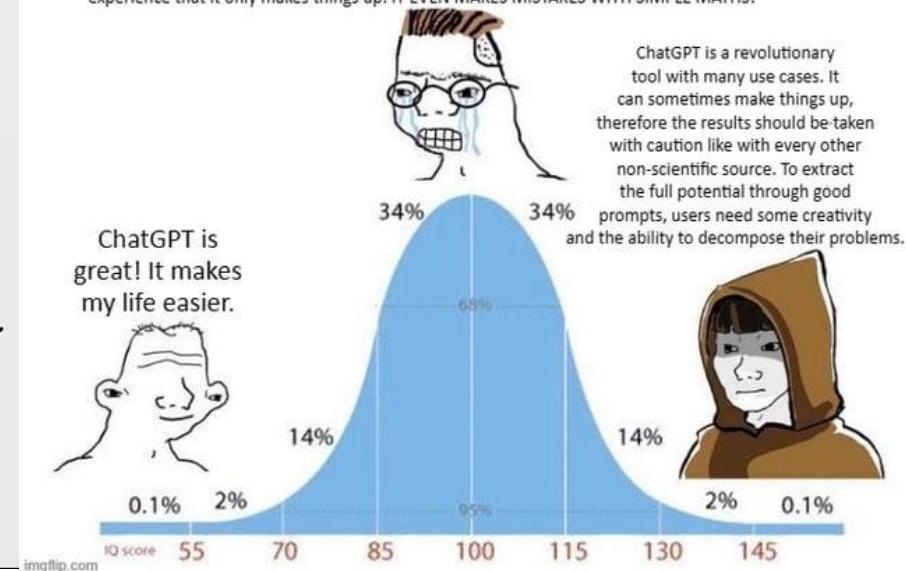
etc,etc...

Evaluate results
beware of
hallucinations!

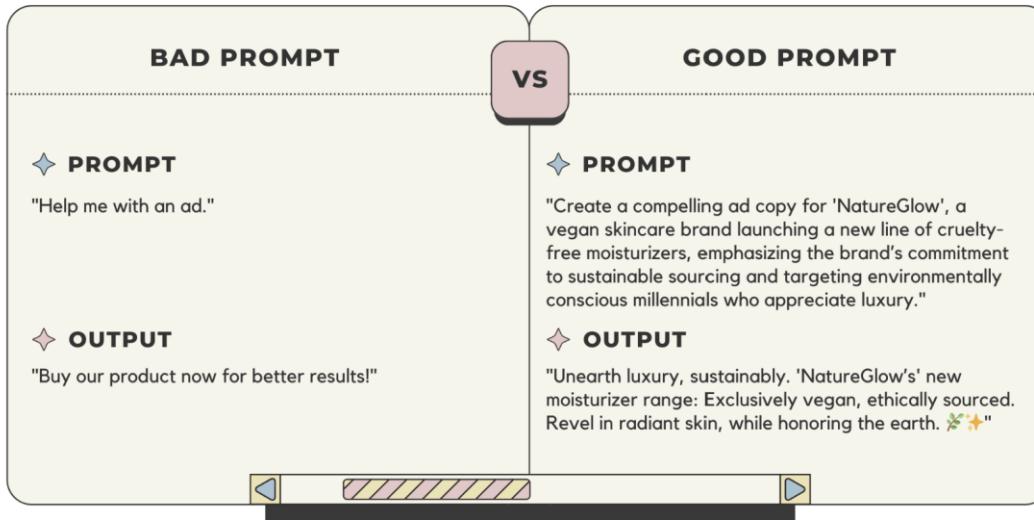
Compensate for AI weaknesses



ChatGPT is overhyped! It's just a LargeLanguageModel and therefore only capable of producing texts. In my highly complex field of work, it is no help at all. When people ask real questions to it, they will experience that it only makes things up. IT EVEN MAKES MISTAKES WITH SIMPLE MATHS!



imgflip.com



GOOD PROMPT	BAD PROMPT
<ul style="list-style-type: none"> • Clear • Specific • Has enough context • Contains audience insights 	<ul style="list-style-type: none"> • Vague • Generic • Has little to no context • Doesn't have any details

Better prompt engineering skills = Better results

Side effect: Better communication skills overall

Study, Practice, Learn

Autonomous Agents with tools



Crew

Agents have the inert ability of reach out to another to delegate work or ask questions



Process

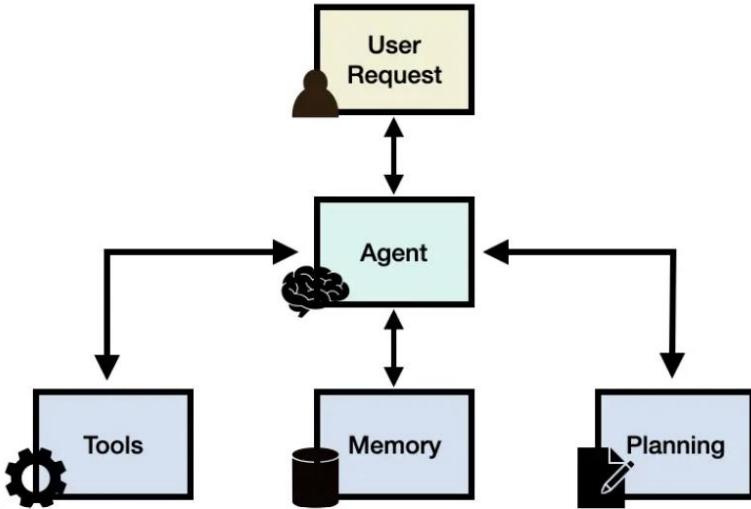
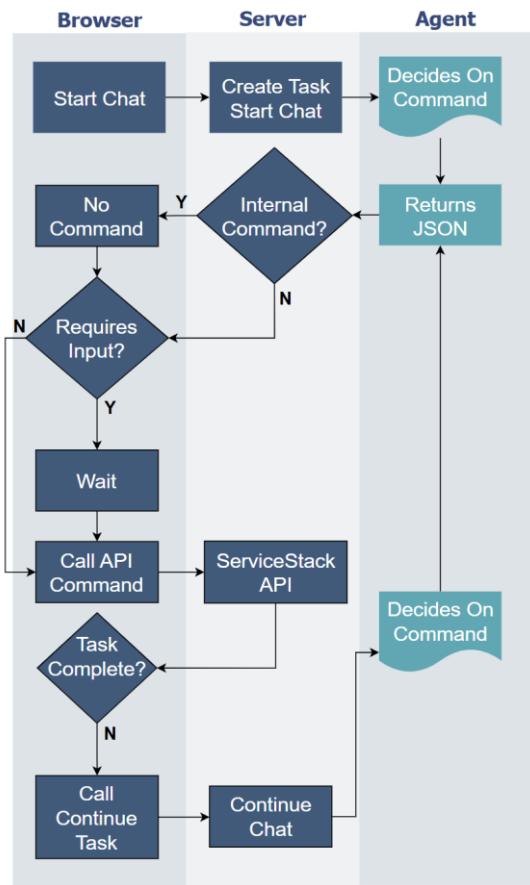


Processes define how agents will work together. How tasks will be assigned. Interaction with each other. How they will perform work.

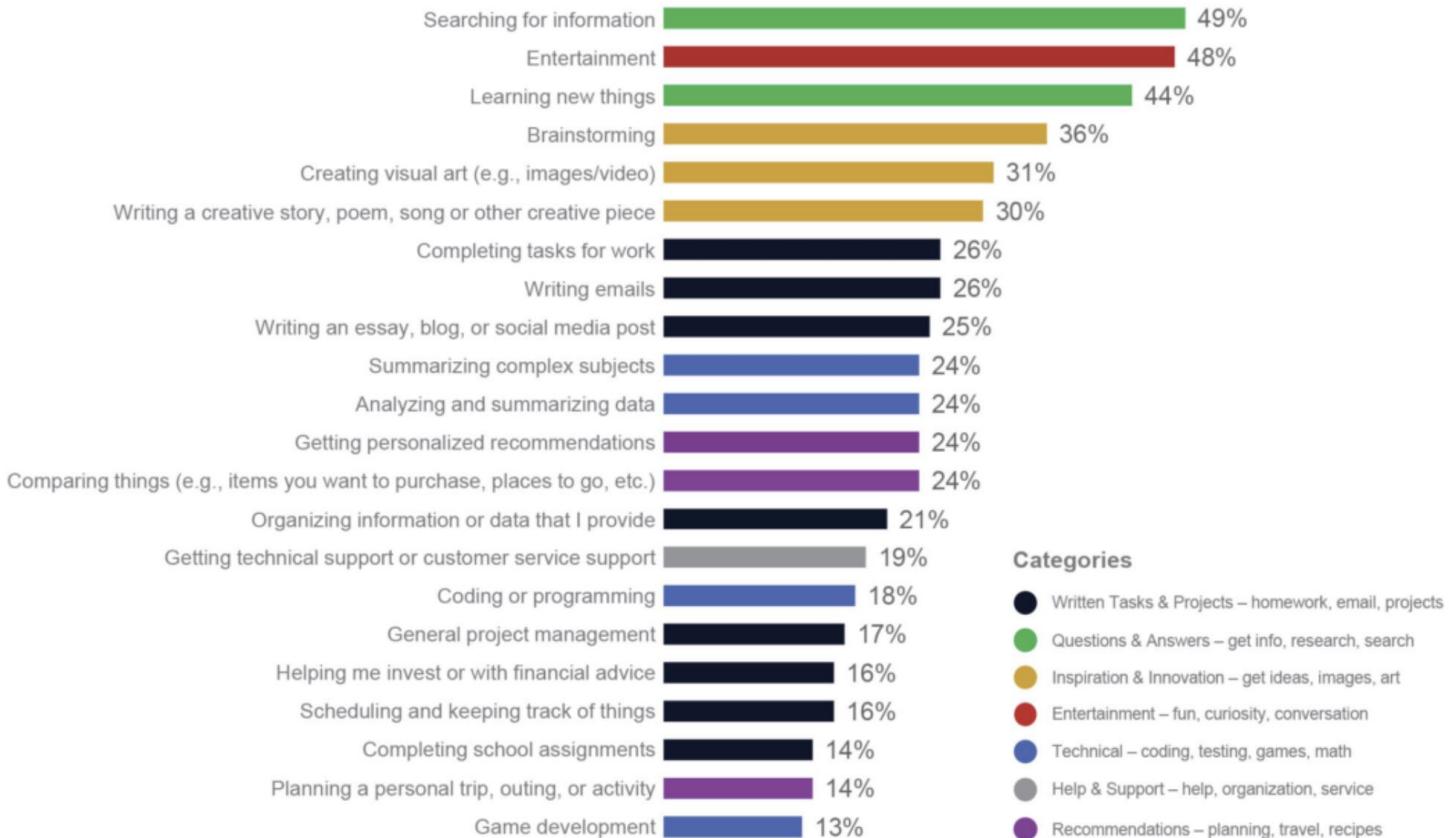


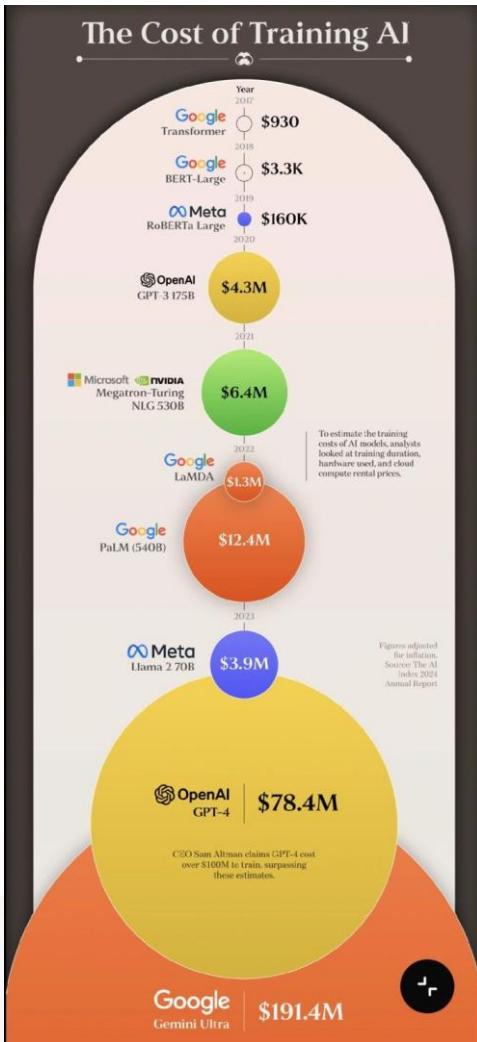
Tasks can override agent tool with specific ones that should be used and also have a specific agent tackle them.

Outcome

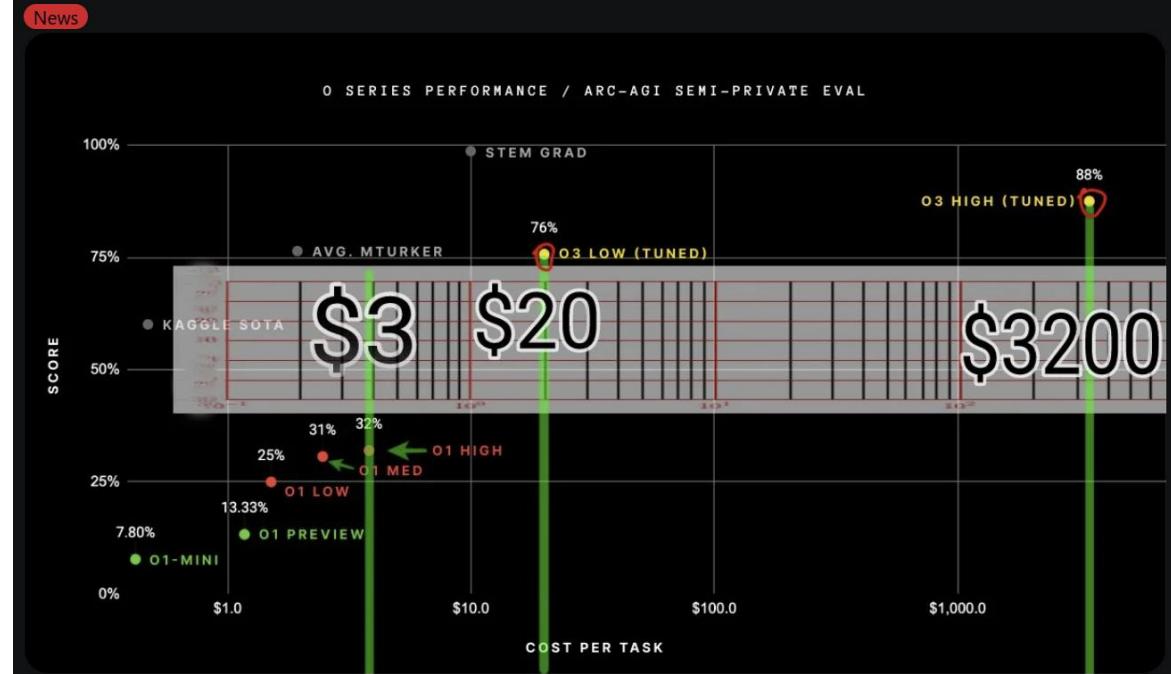


Generative AI Use Case Prevalence





Although OpenAI asked them not to, the cost of O3 was published in this chart.



What Makes Small Language Models so Attractive?

Accessible and Affordable

They can be run (in inference mode) on limited resource regimes (such as laptops and/or small GPUs).



Easier to Customize

Small models can typically be fine-tuned on just a single GPU.

More Energy Efficient

Small language models require fewer computational resources making them more energy-efficient.



Cheaper to Develop

These models only require a relatively small number of GPUs.

Valuable for Educational Purposes

They are more manageable and thus easier to understand and tweak.

Potential Applications of SLMs in Technology and Services



Smartphones



Smart home devices



Wearable Technology



Automotive Systems



Educational Tools



Entertainment Systems

Phi-4 offers high quality results at a small size





Tasks Libraries Datasets Languages Licenses Other

Filter Tasks by name

Multimodal

Text-to-Image Image-to-Text
Text-to-Video Visual Question Answering
Document Question Answering Graph Machine Learning

Computer Vision

Depth Estimation Image Classification
Object Detection Image Segmentation
Image-to-Image Unconditional Image Generation
Video Classification Zero-Shot Image Classification

Natural Language Processing

Text Classification Token Classification
Table Question Answering Question Answering

Models 469,541 Filter by name

meta-llama/Llama-2-70b
Text Generation • Updated 4 days ago • ↓ 25.2k

stabilityai/stable-diffusion-xl-base
Updated 6 days ago • ↓ 2.01k • ❤ 393

openchat/openchat
Text Generation • Updated 2 days ago • ↓ 1.3k • ❤

llyasviel/ControlNet-v1-1
Updated Apr 26 • ❤ 1.87k

cerspense/zeroscope_v2_XL
Updated 3 days ago • ↓ 2.66k • ❤ 334

meta-llama/Llama-2-13b
Text Generation • Updated 4 days ago • ↓ 328 • ❤ 64

USE CASE1:

789012 Bonifico IT60X0542811101000000123456 Mario Rossi Via Roma 123 20100 Milano importo EUR 1500,00 data 05/01/2024
IT60X0542811101000000123456 789012 Via Roma 123 20100
EUR 2300,50 IT72Y0300203280000400145678 Giuseppe Verdi Via Dante 45 50123 Firenze causale: pagamento fattura data 06/01/2024 456789
IT72Y0300203280000400145678 456789 Via Dante 45 50123



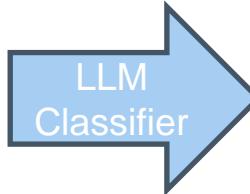
Data	Importo	IBAN	Cod. Cliente	Ordinante	Indirizzo	CAP	Città
05/01/2024	€1.500,00	IT60X0542811101000000123456	789012	Mario Rossi	Via Roma 123	20100	Milano
06/01/2024	€2.300,50	IT72Y0300203280000400145678	456789	Giuseppe Verdi	Via Dante 45	50123	Firenze
07/01/2024	€750,25	IT98Z0336902000001234567890	123456	Luigi Bianchi	Corso Italia 78	30172	Venezia
08/01/2024	€2,00	IT41X0200805364000105917552	987654	Anna Neri	Via Garibaldi 256	00187	Roma

USE CASE2:

Classificazione completata:

	name	year	Price	kms_driven	fuel_type
0	Hyundai Santro Xing XO eRLX Euro III	2007	800.0	45000	Petrol
1	Mahindra Jeep CL550 MDI	2006	4250.0	40	Diesel
2	Hyundai Grand i10 Magna 1.2 Kappa VTG	2014	3250.0	28000	Petrol
3	Ford EcoSport Titanium 1.5L TDCi	2014	5750.0	36000	Diesel
4	Ford Figo	2012	1750.0	41000	Diesel
...
812	Toyota Corolla Altis	2009	3000.0	132000	Petrol
813	Tata Zest XM Diesel	2018	2600.0	27000	Diesel
814	Mahindra Quanto C8	2013	3900.0	40000	Diesel
815	Honda Amaze 1.2 E i VTEC	2014	1800.0	50000	Petrol
816	Chevrolet Sail 1.2 LT ABS	2014	1600.0	50000	Petrol

$$R^2 = 0.23$$



LLM
Classifier

.ficazione completata:

	name	vehicle_type
	Hyundai Santro Xing XO eRLX Euro III	NotClassified
	Mahindra Jeep CL550 MDI	Off-road
	Hyundai Grand i10 Magna 1.2 Kappa VTG	Hatchback
	Ford EcoSport Titanium 1.5L TDCi	SUV
	Ford Figo	Hatchback

	Toyota Corolla Altis	Sedan
	Tata Zest XM Diesel	Sedan
	Mahindra Quanto C8	SUV
	Honda Amaze 1.2 E i VTEC	Sedan
	Chevrolet Sail 1.2 LT ABS	Sedan

$$R^2 = 0.86$$





L'intelligenza artificiale non è solo tecnologia, è uno specchio che ci rimanda domande profonde sulla nostra essenza, sulle scelte che facciamo e su come vogliamo vivere.

Riflessioni per te:

- Se un algoritmo potesse conoserti meglio di te stesso, lasceresti che guidasse le tue decisioni più importanti?
- Come cambierebbero le tue relazioni se un'intelligenza artificiale fosse in grado di comprenderti più a fondo di chiunque altro?

Non si tratta solo di innovare, ma di definire il confine tra ciò che siamo disposti ad affidare alle macchine e ciò che vogliamo preservare come autenticamente umano.

Grazie

youremail@freepik.com

+34 654 321 432

yourwebsite.com



Synthetic data for AI and machine learning is **more flexible** than real data

“Synthetic data generation accelerates the analytics development cycle, lessens regulatory concerns and lowers the cost of data acquisition.” – Gartner

Synthetic data is the perfect fuel that AI and machine learning development projects need. Synthetic data is created by generative AI algorithms, which can be instructed to create bigger, smaller, fairer or richer versions of the original data. Due to how the synthesization process takes place, the data can be augmented to fit certain characteristics. In a way, synthetic data is like modelling clay for data scientists and data managers. For example, upsampling minority groups in a dataset can improve the performance of machine learning models. Or human bias embedded in the original data can be removed by introducing fairness constraints to the generation process.

Value of Synthetic Data

- **Model Training:** Provides vast training data with tailored scenarios for robust models.
- **Data Sharing:** Enables collaboration without privacy concerns.
- **Bias Mitigation:** Addresses biases in real data for fairer applications.
- **Testing and Validation:** Supports comprehensive testing in controlled environments.
- **Cost Reduction:** Reduces data collection and labeling costs.



Real life examples of synthetic data projects

AI-generated synthetic data is used by companies across the world to accelerate their data innovation projects in a privacy-compliant and agile way.

- ▶ Telefónica uses synthetic customer data for analytics
- ▶ Erste Bank used synthetic test data to develop a successful mobile banking app
- ▶ JPMorgan leverages a synthetic data sandbox to speed up data-intensive POCs with third party vendors
- ▶ Anthem uses synthetic data to detect fraud and for delivering personalized service to members

Method	Protection from re-identification risk	Feature Statistics	Feature correlations	ML performance
Synthetic data	High	High	High	High
Randomization	Low	Medium	Low	Low
Permutation	Low	High	Very low	Very low
Generalization	Medium	Low	Low	Low
Pseudonymization	Very low	High	High	High
Data masking	Very low	Very low	Very low	Very low

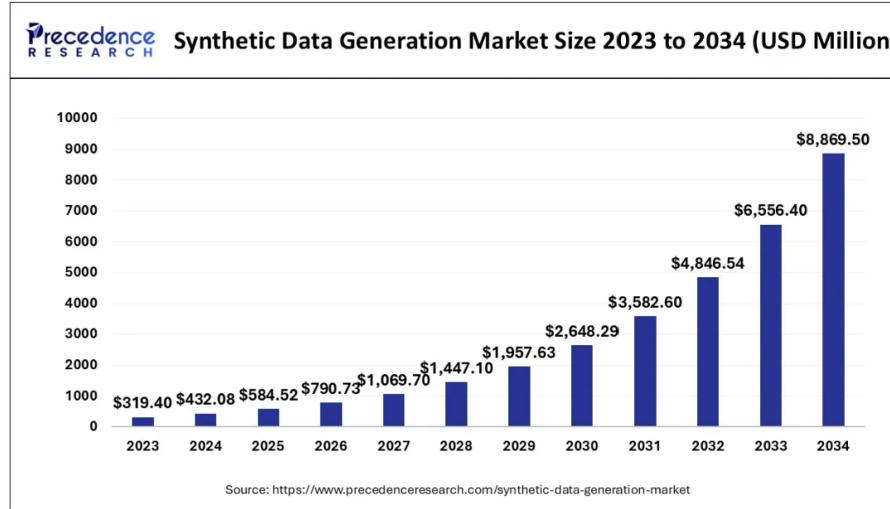
How does synthetic data compare to other data anonymization techniques?

Research has shown many times that legacy data anonymization techniques endanger privacy. They often also degrade the utility of the data.

Even worse – a lot of companies are equating pseudonymization with anonymization. But from a legal perspective, pseudonymized data is still personal data. And it needs to be treated and protected as just that.

How can synthetic data help data-driven businesses?

Extract the full value of your data assets. We enable responsible privacy enhancing data availability, free of bias, and incompleteness.



DIPENDENZA EMOTIVA E GHOSTING: I RISCHI DELL'UTILIZZO DI CHATGPT

OpenAI mette in guardia rispetto a possibile "antropomorfizzazione e dipendenza emotiva" dall'IA

OpenAI ha messo in guardia gli utenti rispetto ad alcuni **potenziali effetti negativi dell'uso di ChatGPT**, presentati nel rapporto '**GPT-4o System Card**', dedicato a tutte le analisi di sicurezza svolte prima del rilascio di **GPT-4o**.

In particolare, i **rischi** che preoccupano maggiormente riguardano **gli effetti di un'interazione prolungata con ChatGPT**:



la possibilità che insegni agli utenti comportamenti negativi nelle relazioni interpersonali, come il ghosting



l'eventualità di portare l'utente ad una sorta di dipendenza emotiva

Perché potrebbe verificarsi tutto questo? L'intelligenza artificiale è una tecnologia estremamente innovativa, in rapida e costante evoluzione. Le sue applicazioni portano **enormi benefici, ma anche rischi da non sottovalutare.** Riuscire a **garantire** che venga **utilizzata in modo etico e responsabile** è fondamentale.

È per questo che **OpenAI** sta **collaborando con sociologi e psicologi** mentre **intensifica le sue ricerche e sviluppa nuove strategie di sicurezza.**

commentary in snack-sized mouthfuls



"I want AI to do my laundry and dishes so that I can do art and writing, not for AI to do my art and writing so that I can do my laundry and dishes."

Author and videogame enthusiast **Joanna Maciejewska** nails it
(although bathroom cleaning goes ahead of laundry and dishes)

"I'm sure I deserve
a lot of ..."