

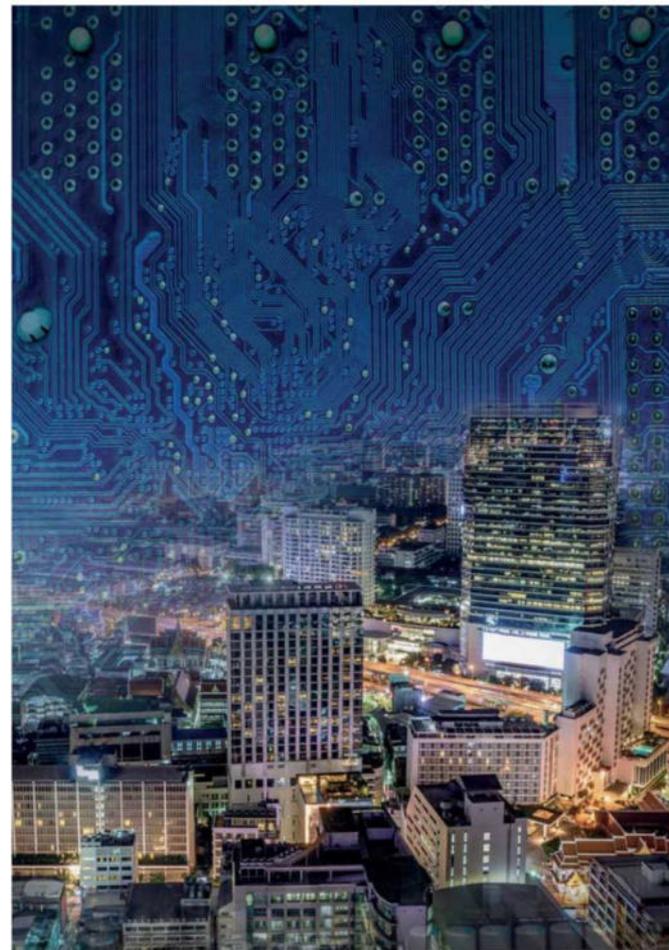
Introduzione all'intelligenza artificiale

13 giugno 2023

Andrea Bernardini

Argomenti del giorno

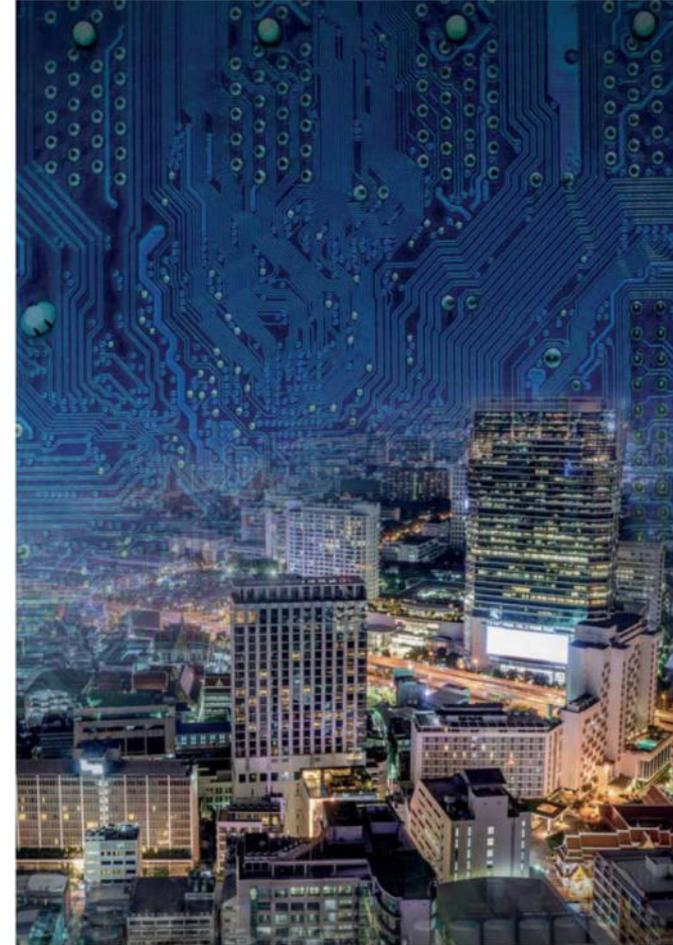
- Introduzione
- Strumen. di AI nel 2023
- Cenni storici e terminologia
- Modelli di Intelligenza Ar.ficiale
- Nuove evoluzioni dell'AI
- Norma.va e a?vità della UE



Argomenti del giorno

- Introduzione
- Strumen. di AI nel 2023
- Cenni storici e terminologia
- Modelli di Intelligenza Ar.ficiale
- Nuove evoluzioni dell'AI
- Norma.va e a?vità della UE

Ma prima di cominciare qualche domanda



BT to cut 55,000 jobs with up to a fifth replaced by AI

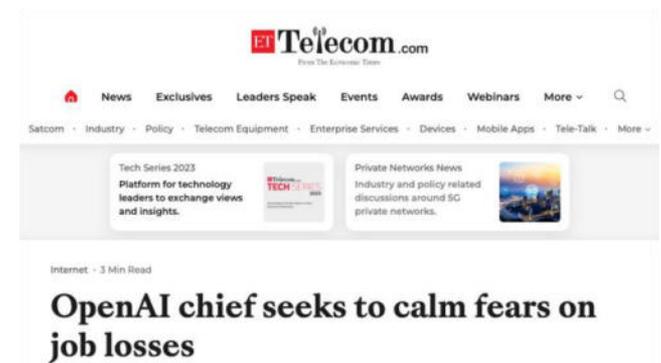
18 May · Comments



BT GROUP

By Tom Espiner
Business reporter, BBC News

Automazione e IA rimpiazzeranno molti posti di lavoro perché le reti 5G avranno bisogno di molta meno manodopera delle attuali reti di telecomunicazione, meno interventi e meno controlli, tutto o quasi sarà effettuato da remoto e comunque serviranno molti meno dipendenti ha spiegato il CEO di BT, Philip Jansen.



Fonte:

<https://www.bbc.com/news/business-65631168>

<https://www.key4biz.it/bt-tagliera-10-mila-posti-di-lavoro-per-far-posto-allia-entro-il-2030/446818/>

<https://telecom.economictimes.indiatimes.com/news/internet/openai-chief-seeks-to-calm-fears-on-job-losses/100540898>

E in Germania ChatGpt celebra la messa: "È stata un'omelia solida ma senza cuore"

di Natasha Caragnano



Nella città bavarese di Fuerth in centinaia hanno partecipato alla funzione religiosa, organizzata in occasione del Congresso della Chiesa Evangelica tedesca. Il testo è stato scritto al 98% dall'Intelligenza artificiale e predicata da 4 avatar

Fonte: https://www.repubblica.it/esteri/2023/06/10/news/chatgpt_celebra_messa_germania-403947969/

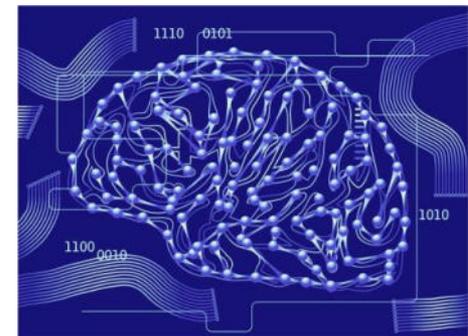
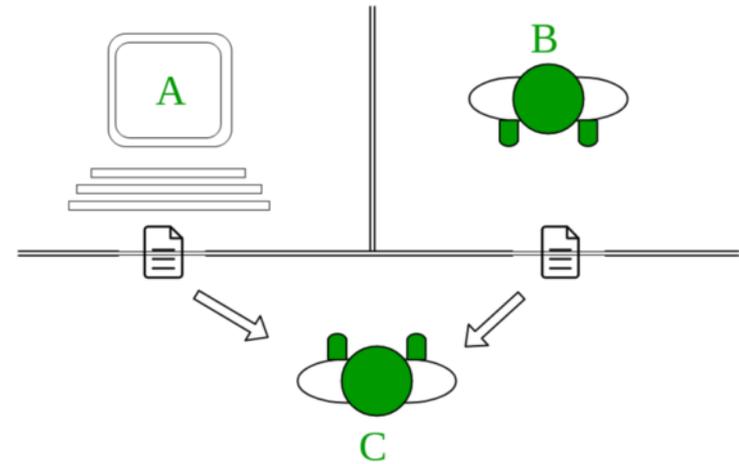
INTRODUZIONE

Intro - Definizione semplice di IA

“L’intelligenza artificiale permette la progettazione di sistemi sia hardware sia software in grado di dotare le macchine di caratteristiche che vengono considerate tipicamente umane”.

Teorizzazione del Turing test (Alan Turing 1950), in cui il giocatore C, l'interrogatore, ha il compito di provare a determinare quale giocatore - A o B - è un computer e quale è un umano. L'interrogatore si limita a utilizzare le risposte alle domande scritte per effettuare la determinazione. Un computer passa il test, se può ingannare un interrogatore umano.

Componenti principali della IA secondo Turing: (linguaggio, ragionamento, conoscenza, apprendimento e comprensione).



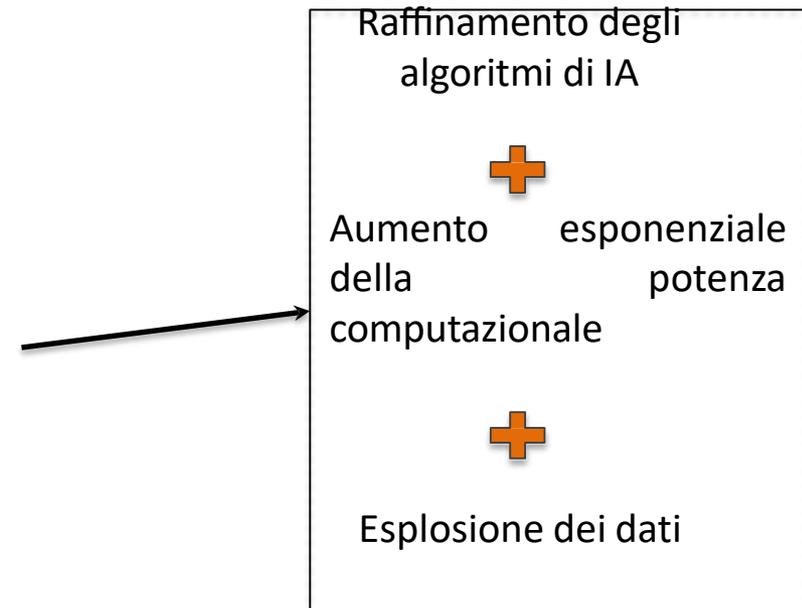
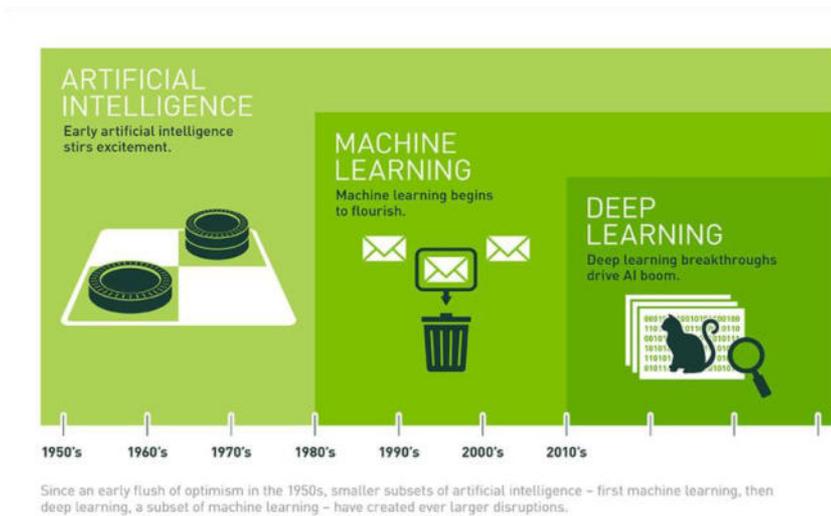


L'Intelligenza artificiale (IA) concerne i sistemi che manifestano un comportamento intelligente nell'analizzare il loro ambiente e nell'intraprendere azioni, con un certo grado di autonomia, per conseguire determinati obiettivi.

I sistemi di 'IA' possono essere basati sul solo software e agire nel **dominio virtuale** (assistenti vocali, software per l'analisi delle immagini, motori di ricerca, sistemi per il riconoscimento della voce e dei volti), oppure possono far parte di **dispositivi hardware** (robot avanzati, automobili autonome, droni e internet delle cose).

Fonte: <https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines>, April 2019

Le origini dell'IA (Intelligenza Artificiale) risalgono agli anni '40 e '50, ma è solo di recente entrata a far parte del nostro quotidiano grazie ad applicazioni di filtraggio dello spam, traduzione automatica, riconoscimento e generazione di oggetti multimediali, sistemi di dialogo intelligenti e molte altre applicazioni. Lo sviluppo delle reti di sensori e dell'IoT sta aumentando a dismisura le possibilità di percezione dell'ambiente della IA, il volume dei dati da analizzare e l'impatto sulla società delle decisioni prese dai sistemi di IA.



Intelligenza Artificiale debole (weak/specialist IA): racchiude al suo interno sistemi in grado di simulare alcune funzionalità cognitive dell'uomo senza tuttavia raggiungere le capacità intellettuali tipiche dell'uomo; si tratta, a grandi linee, di programmi di problem-solving in grado di replicare alcuni ragionamenti logici umani per risolvere problemi, prendere decisioni, ecc.

Intelligenza Artificiale forte (strong/general IA): si fanno rientrare in questa categoria i sistemi in grado di diventare sapienti (o addirittura coscienti di sé); ci sono teorie che spingono alcuni scienziati ed esperti a ritenere che un giorno le macchine avranno una intelligenza propria, autonoma e probabilmente superiore a quella degli esseri umani.

“The development of full artificial intelligence could spell the end of the human race.” Stephen Hawking



A specialist A.I. can drive a car or play chess, not very good outside it's domain.



General A.I. would have a human or superhuman level of intelligence

Settori di utilizzo della AI (debole)

Speech recognition

Autonomous planning and
scheduling

Financial forecasting

Spam fighting

Logistics planning

Robotics (household, surgery, navigation)

Machine translation

Information extraction

Automatic assembly

Sentiment analysis

Computer animation

Fraud detection

Recommendation systems

Web search engines

Autonomous cars

Energy optimization

Question answering systems

Social network analysis

Medical diagnosis

Route finding

Protein design

Document summarization

Transportation/scheduling

Google

riconoscimento vocale e parlato, classificazione email, suggeritore di risposte per email, calcolo di percorsi nelle mappe....

Facebook

Classificazione e personalizzazione dei contenuti, filtraggio di contenuti offensivi, riconoscimento contenuti immagini e video...

Amazon

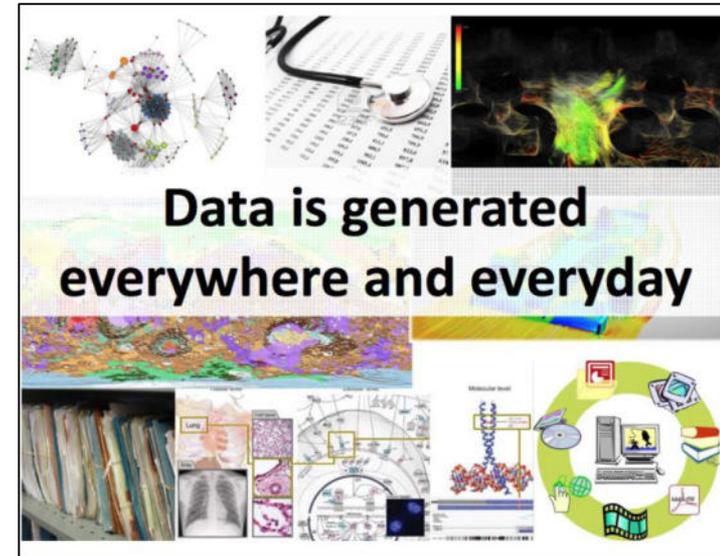
Raccomandazioni di prodotti e gestione della rete di distribuzione...

LinkedIn

Personalizzazione, comprensione del linguaggio, sistemi di raccomandazioni...

E tantissimi altri servizi

Ride-sharing apps, smart assistant, etc..



Google

riconoscimento volti e parlato, classificazione email, suggeritore di risposte per mail, calcoli di percorsi nelle mappe...

Facebook

Classificazione e personalizzazione dei contenuti, filtraggio di contenuti offensivi, riconoscimento contenuti immagini e video...

Amazon

Raccomandazioni di prodotti e gestione della rete di distribuzione...

Linkedin

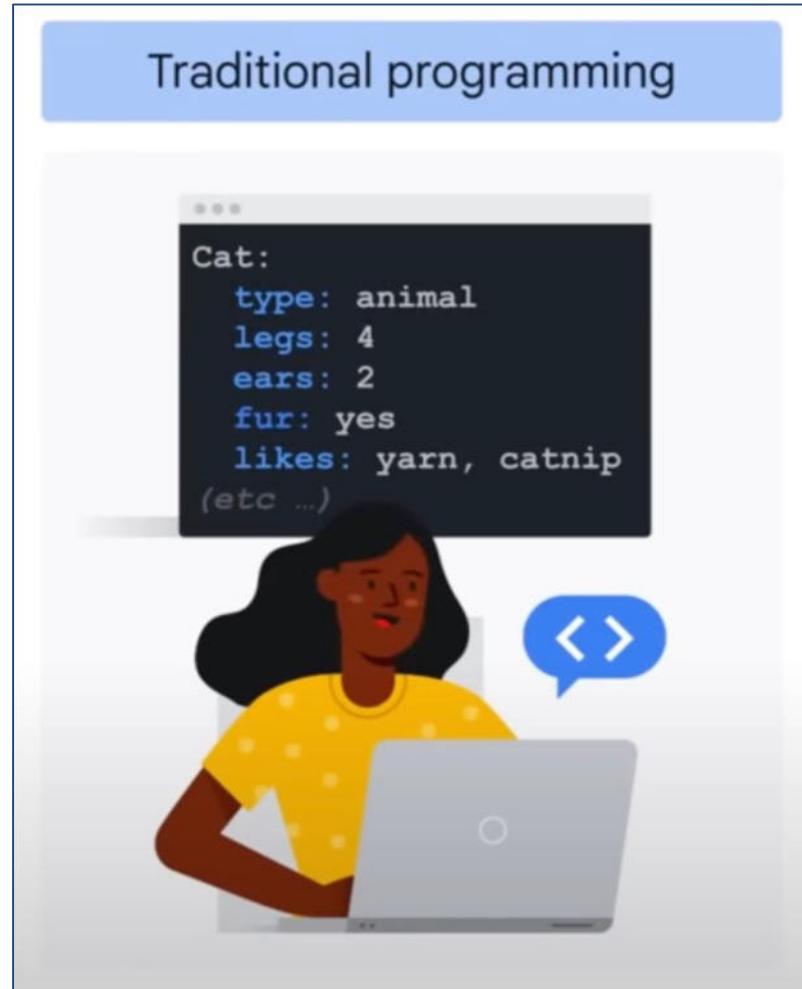
Personalizzazione, comprensione del linguaggio, sistemi di raccomandazioni...

E tantissimi altri servizi

Ride-sharing apps, smart assistant, etc..

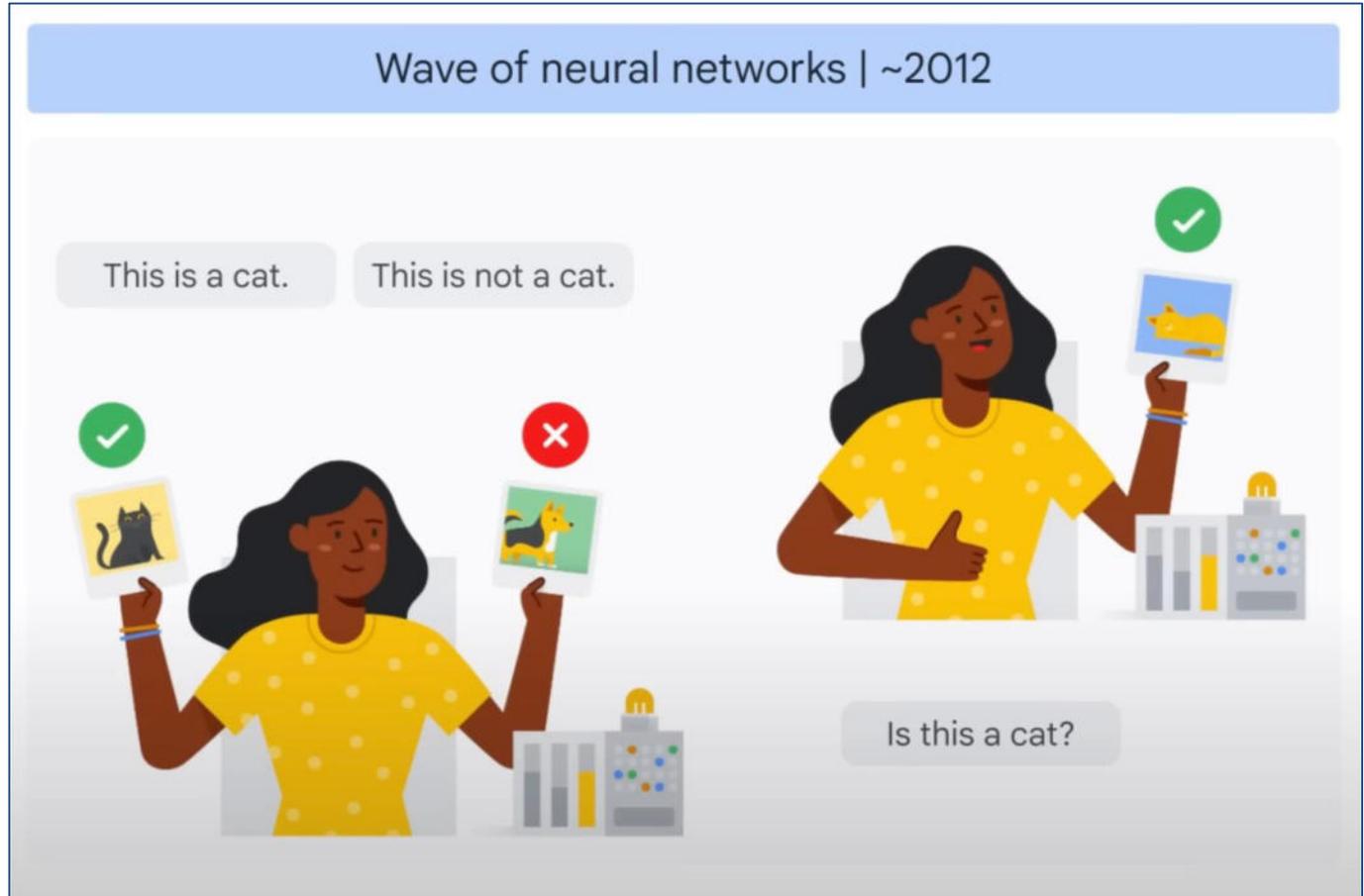
Modelli di linguaggio di grandi dimensioni (LLM) e la AI generativa (ChatGPT, Bard,...)





Fonte: Google Cloud Tech

Il ritorno delle reti neurali

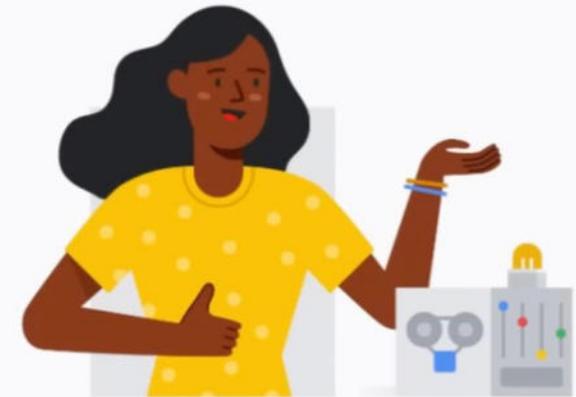
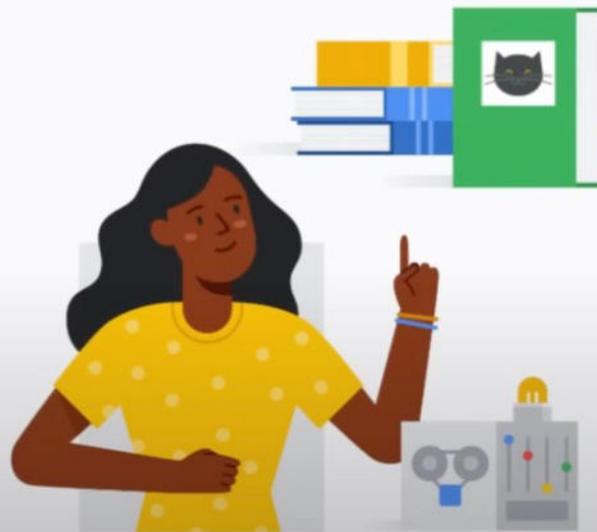


Fonte: Google Cloud Tech

I modelli di linguaggio generativi

Generative language models | LaMDA, PaLM, GPT, etc.

Go read this huge pile of books.



So, you've learned about cats and millions of other concepts

What's a cat?

A cat is ...

Fonte: Google Cloud Tech

Deep Learning Model Types



Discriminative

- Used to classify or predict
- Typically trained on a dataset of labeled data
- Learns the relationship between the features of the data points and the labels

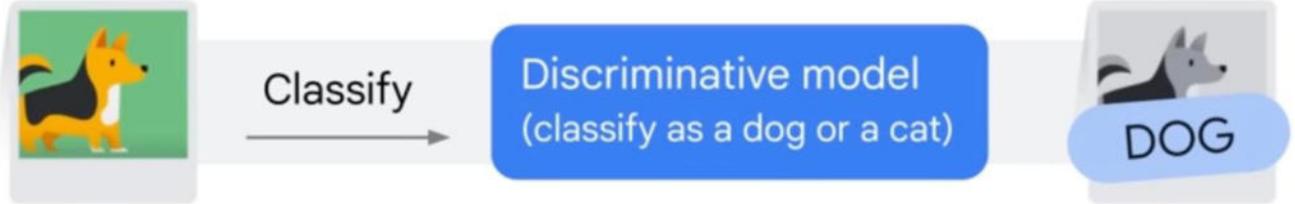


Generative

- Generates new data that is similar to data it was trained on
- Understands distribution of data and how likely a given example is
- Predict next word in a sequence

Dai modelli discriminativi a quelli generativi

Discriminative
technique



Generative
technique



Fonte: Google Cloud Tech

ChatGPT è un chatbot basato su intelligenza artificiale e apprendimento automatico sviluppato da OpenAI specializzato nella conversazione con un utente umano. La sigla GPT sta per **Generative Pre-trained Transformer**, una nuova tecnologia applicata al machine learning.

ChatGPT è stato messo a punto con tecniche di machine learning non supervisionato e ottimizzato con tecniche di apprendimento supervisionato e per rinforzo.

ChatGPT software	
	
	
Genere	Chatterbot
Sviluppatore	OpenAI
Data prima versione	30 novembre 2022; 5 mesi fa
Ultima versione	12 maggio 2023; 14 giorni fa ^[1]
Sistema operativo	nessuno
Linguaggio	Python
Licenza	Software proprietario (licenza non libera)
Sito web	chat.openai.com/ 
<small>Modifica dati su Wikidata · Manuale</small>	

ChatGPT è un chatbot basato su intelligenza artificiale e apprendimento automatico sviluppato da OpenAI specializzato nella conversazione con un utente umano. La sigla GPT sta per **Generative Pre-trained Transformer**, una nuova tecnologia applicata al machine learning.

ChatGPT è stato messo a punto con **tecniche di machine learning non supervisionato e ottimizzato con tecniche di apprendimento supervisionato e per rinforzo.**

ChatGPT software	
	
	
Genere	Chatterbot
Sviluppatore	OpenAI
Data prima versione	30 novembre 2022; 5 mesi fa
Ultima versione	12 maggio 2023; 14 giorni fa ^[1]
Sistema operativo	nessuno
Linguaggio	Python
Licenza	Software proprietario (licenza non libera)
Sito web	chat.openai.com/ 
<small>Modifica dati su Wikidata · Manuale</small>	

Addestramento

Il modello è stato addestrato su miliardi di campioni di testo da Internet (oltre 175 miliardi di parametri per GPT-3) 30 trilioni di parole da pagine web, libri, articoli di notizie, forum, testi scientifici e molte altre fonti di informazioni pubbliche disponibili fino a settembre 2021.

Utilizzo

I modelli di linguaggio generativi possono scrivere testo originale in base alle richieste dell'utente. Il testo è grammaticalmente corretto e i paragrafi sono ben strutturati.

Alcune attività possibili sono: generare idee e rispondere a domande, scrivere saggi, scrivere opere creative come poesie e canzoni, produzione di testi anche tecnici, scrivere codice di programmazione

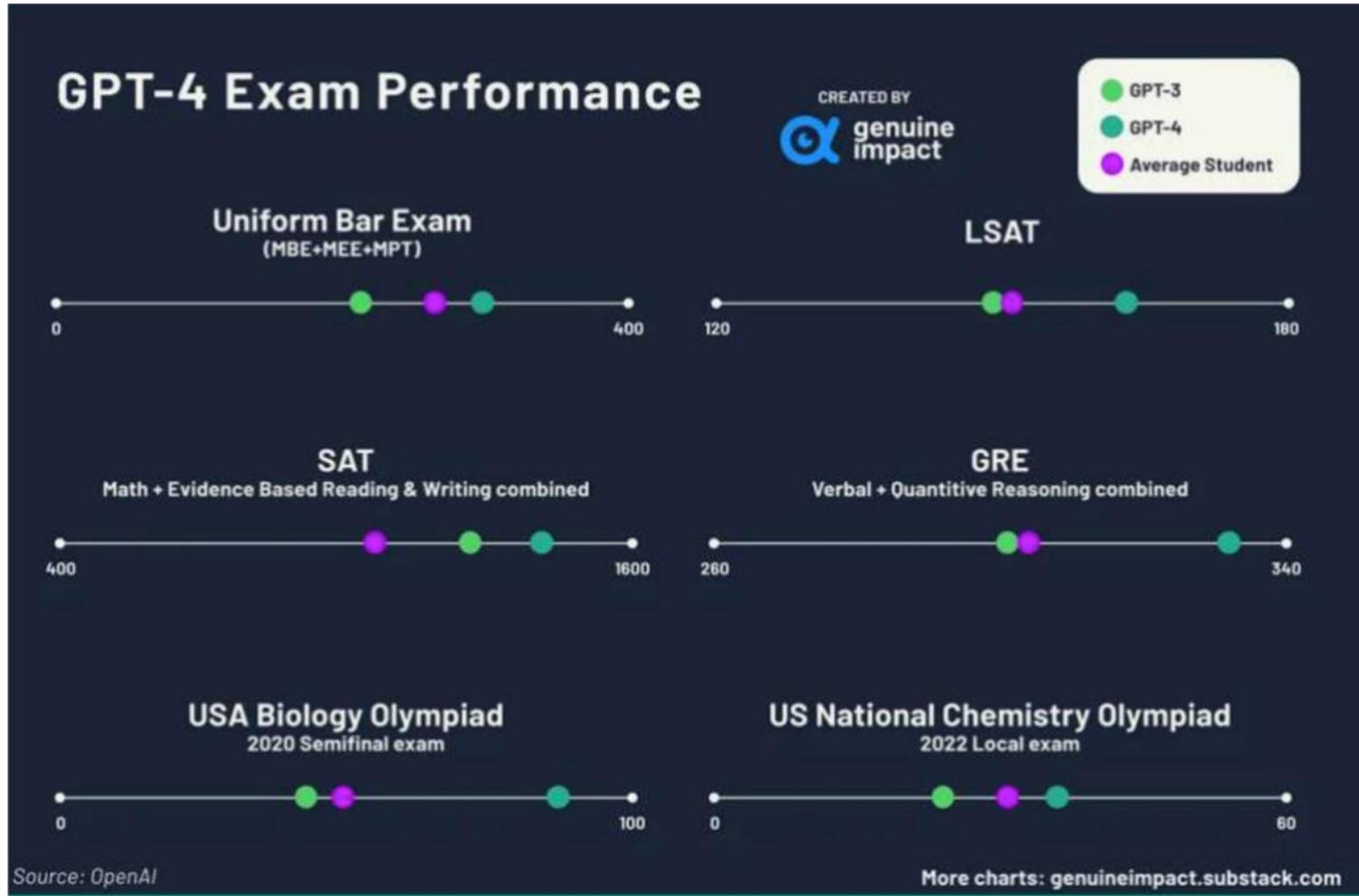
- Padronanza di linguaggio è un effetto collaterale di predizione delle parole di un testo che ha sorpreso anche gli sviluppatori. In sostanza, fa un ottimo lavoro nel prevedere ciò che un essere umano potrebbe scrivere
- Si basa su modelli statistici, non hanno la capacità di comprendere il testo che viene generato e non determina se le informazioni sono vere o false
- Presto per capire il legame con la vera intelligenza
- Noam Chomsky dice che ingegneristicamente interessante ma niente di più

Fon4: Weidinger, et al., 2022, Ethical and social risks of harm from Language Models

Noam Chomsky: The False Promise of ChatGPT,

<https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>

Capacità di superare test ed esami



Come usarlo ? (1) - Scrivere documenti

A

Sono un avvocato vorrei che mi scrivessi un atto di citazione per risarcimento per danni che un mio cliente ha avuto in casa da una perdita nel tubo del suo vicino di casa del piano superiore.



Sono un assistente virtuale e posso fornirti un esempio di bozza di atto di citazione per risarcimento danni causati da una perdita nel tubo del vicino di casa del piano superiore. Ricorda che questo è solo un esempio generico e potrebbe richiedere personalizzazioni specifiche in base alle leggi e alle normative del tuo paese o della tua giurisdizione. È sempre consigliabile consultare un avvocato professionista per ricevere assistenza legale adeguata. Ecco un esempio di bozza di atto di citazione:

[Nome del tuo Studio Legale]
[Indirizzo del tuo Studio Legale]
[Numero di telefono]
[Indirizzo email]

Tribunale di [Nome del Tribunale]
Causa n. [Numero di Causa]

[Il tuo Cliente], ricorrente,
C.F./P.IVA: [Codice Fiscale o Partita IVA del tuo Cliente],
Residente a [Indirizzo di Residenza del tuo Cliente],
Città di [Città di Residenza del tuo Cliente],
Telefono: [Numero di Telefono del tuo Cliente],
Email: [Indirizzo Email del tuo Cliente],

contro

[Nome del Vicino], convenuto,
C.F./P.IVA: [Codice Fiscale o Partita IVA del Vicino],
Residente a [Indirizzo del Vicino],
Città di [Città del Vicino].

OGGETTO: ATTO DI CITAZIONE PER RISARCIMENTO DANNEGGIAMENTO

Premesse:

1. Il ricorrente, [Il tuo Cliente], è proprietario e residente presso l'immobile sito in [Indirizzo della Proprietà del tuo Cliente].
2. Il convenuto, [Nome del Vicino], è proprietario e residente presso l'immobile del piano superiore sito in [Indirizzo del Vicino].

Fatti:

1. Il [Data dell'incidente], si è verificata una perdita nel tubo idraulico dell'immobile del convenuto, situato al piano superiore, che ha causato danni significativi all'immobile del ricorrente.
2. La perdita è stata causata dalla negligenza e dalla mancanza di manutenzione adeguata da parte del convenuto.
3. A causa di tale perdita, si è verificato un danneggiamento dell'immobile del ricorrente, includendo ma non limitandosi a infiltrazioni d'acqua, macchie sulle pareti, danni strutturali e costi di riparazione e ripristino.
4. Il convenuto è responsabile per i danni subiti dal ricorrente in virtù della sua negligenza nell'adottare misure correttive tempestive.

Richiesta:

In base a quanto sopra, si richiede al Tribunale di:

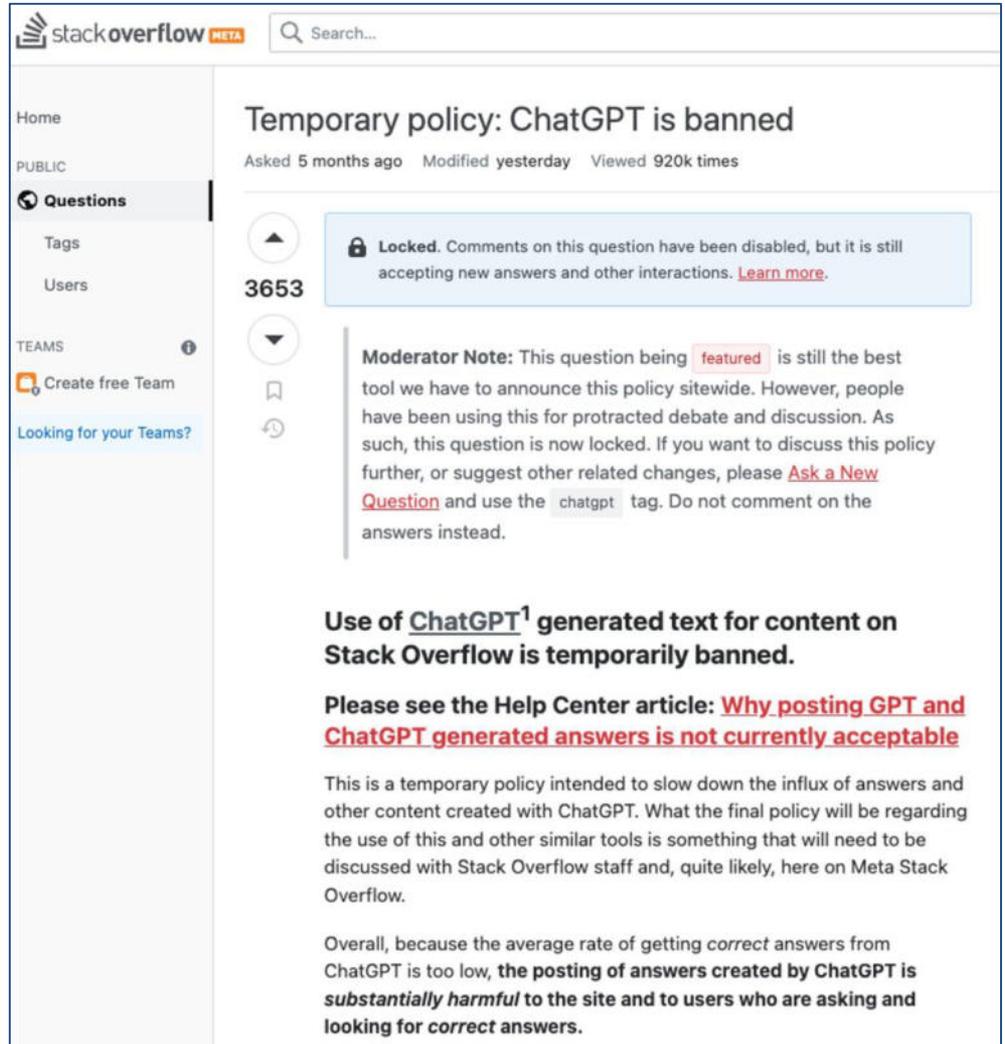
1. Condannare il convenuto a risarcire integralmente il ricorrente per i danni subiti, pari a [importo richiesto in cifre e lettere], a titolo di risarcimento per i danneggiamenti all'immobile del ricorrente causati dalla perdita nel tubo idraulico dell'immobile del convenuto del piano superiore.
2. Condannare il convenuto a rimborsare al ricorrente tutte le spese sostenute per

Idea da: <https://www.altalex.com/documents/news/2023/01/16/la-chatgpt-che-scriveva-come-un-avvocato>

“Nel passaggio “evolutivo” della ChatGPT pare scomparsa la possibilità di richiudere la redazione di atti giudiziari veri e propri, a meno che tu non specifichi di essere un legale: in qualche modo la chatbot deve essere stata ben istruita sull'intervento necessario degli avvocati...”

The primary problem is that **while the answers which ChatGPT produces have a high rate of being incorrect, they typically look like they might be good and the answers are very easy to produce.** There are also many people trying out ChatGPT to create answers, without the expertise or willingness to verify that the answer is correct prior to posting.

Fonte:
<https://meta.stackoverflow.com/questions/421831/temporary-policy-chatgpt-is-banned>



The screenshot shows a Stack Overflow Meta page with the title "Temporary policy: ChatGPT is banned". The post is 5 months old, modified yesterday, and has been viewed 920k times. It has 3653 votes. A blue lock icon indicates that comments are disabled. The post content includes a "Moderator Note" explaining that the question is locked because ChatGPT-generated answers are not currently acceptable. The note states that the question is still the best tool to announce this policy, but people are using it for protracted debate. It asks users to "Ask a New Question" with the "chatgpt" tag and not to comment on answers. Below the note, there is a bolded statement: "Use of ChatGPT¹ generated text for content on Stack Overflow is temporarily banned." followed by a link to a Help Center article: "Why posting GPT and ChatGPT generated answers is not currently acceptable". The article text explains that this is a temporary policy to slow down the influx of answers and content created with ChatGPT, and that the final policy will be discussed with Stack Overflow staff. It concludes that the average rate of getting correct answers from ChatGPT is too low, and that the posting of answers created by ChatGPT is substantially harmful to the site and to users who are asking and looking for correct answers.

The screenshot displays the GitHub Copilot interface within a code editor. On the left, the Copilot chat window is open, showing a welcome message to @martinwoodward and several suggested actions: 'Generate unit tests for my code.', 'Explain the selected code.', and 'Propose a fix for the bugs in my code.' Below these are links to documentation and a search bar with the placeholder text 'Ask a question or type '/' for topics'. The main editor window shows a Python file named 'add_elements.py' with a function 'parse_expenses' that processes a string of expense data. The code includes comments and a list of tuples representing expenses. At the bottom, the terminal window shows a 'ValueError: not enough values to unpack (expected 3, got 1)' occurring at line 17 of the file, with a dropdown menu showing 'Python' as the selected language.

Fonte: https://github.com/github-copilot/chat_waitlist_signup

Come usarlo ? (4) - Integrazione di API nel codice di programmi

```
import pandas as pd
from pandasai import PandasAI

# Sample DataFrame
df = pd.DataFrame({
    "country": ["United States", "United Kingdom", "France", "Germany", "Italy", "Spain", "Canada",
    "gdp": [19294482071552, 2891615567872, 2411255037952, 3435817336832, 1745433788416, 118120513536,
    "happiness_index": [6.94, 7.16, 6.66, 7.07, 6.38, 6.4, 7.23, 7.22, 5.87, 5.12]
})

# Instantiate a LLM
from pandasai.llm.openai import OpenAI
llm = OpenAI(api_token="YOUR_API_TOKEN")

pandas_ai = PandasAI(llm)
pandas_ai(df, prompt='Which are the 5 happiest countries?')
```

```
6         Canada
7         Australia
1    United Kingdom
3           Germany
0    United States
Name: country, dtype: object
```

Fonte: <https://github.com/gventuri/pandas-ai>

Come usarlo ? (5) - Cheat sheet




The ChatGPT Cheat Sheet

Act as a [ROLE]

1. Analyst
2. Teacher
3. Marketer
4. Advertiser
5. Mindset Coach
6. Therapist
7. Journalist
8. Inventor
9. Lawyer
10. Ghostwriter
11. Website Designer
12. Best Selling Author
13. Chief Financial Officer
14. Expert Copywriter
15. Prompt Engineer
16. Accountant
17. Project Manager
18. Sports Coach
19. Financial Analyst
20. Full Stack Developer
21. Linux Terminal
22. "Position" Interviewer
23. Excel Sheet
24. Plagiarism Checker
25. Relationship Coach
26. Recruiter
27. JavaScript Console

Create a [TASK]

1. Headline
2. Article
3. Essay
4. Book Outline
5. Email Sequence
6. Social Media Post
7. Product Description
8. Cover Letter
9. Blog Post
10. SEO Keywords
11. Summary
12. Video Script
13. Recipe
14. Sales Copy
15. Analysis
16. Ad Copy
17. Web Page

Show as [FORMAT]

1. A Table
2. A List
3. Summary
4. HTML
5. Code
6. Spreadsheet
7. Graphs
8. CSV file
9. Plain Text file
10. JSON
11. Rich Text
12. PDF
13. XML
14. Markdown
15. Gantt Chart
16. Word Cloud
17. Emojis
18. An Analogy
19. Bullet Points

Set restrictions

1. Use poetic language
2. Adopt a formal tone
3. Write short sentences
4. Code only in HTML/CSS
5. Use Shakespearian style
6. Write using basic English
7. Use only scientific sources
8. Add pop culture references

Prompts For Biz Owners

1. Give me a list of inexpensive ideas on how to promote my business better.
2. Acting as a Business Consultant, What is the best way to solve this problem of [Problem]?
3. Create a 30-Day Social Media Content Strategy based on [Topic 1] & [Topic 2].

Prompts for Developers

1. Develop an architecture and code for a <description> website with JavaScript.
2. Help me find mistakes in the following code <paste code below>
3. I want to implement a sticky header on my website. Can you provide an example using CSS and JavaScript?
4. Please continue writing this code for JavaScript <past code below>

Prompts for Marketers

1. Can you provide me with some ideas for blog posts about [topic]?
2. Write a product description for my [product or service or company].
3. Suggest inexpensive ways I can promote my [company] without using social media.
4. How can I obtain high-quality backlinks to improve the SEO of [website name]?

Prompt Priming

1. ZERO - "Write me 5 Headlines about [Topic]"
2. SINGLE - "Write me 5 Headlines about [Topic]. Here is an example of one headline: 5 Ways to Lose Weight"
3. MULTIPLE - "Write me 5 Headlines about [Topic]. Here is an example of some headlines: 5 Ways to Lose Weight, How to Lose More Fat in 4 Weeks, Say Goodbye to Stubborn Fat, Find a faster way to Lose Weight Fast"

Prompts for Designers

1. Generate examples of UI design requirements for a [mobile app].
2. How can I design a [law firm website] in a way that conveys [trust and authority]?
3. What are some micro-interactions to consider when designing a Fintech app?
4. Create a text-based Excel sheet to input your copy suggestions. Assume you have 3 members in your UX writing team.

C.R.E.A.T.E Formula for Prompt Engineering

1. **C: Character** - Define the AI's role. E.g., "You are a seasoned copywriter with 20 years of experience in successful sales copy."
2. **R: Request** - Be specific with what you need. Instead of "Write a sales email for a sports car," specify, "Write a compelling email for the Cougar Hyper Sport, an electric car with top-tier acceleration."
3. **E: Examples** - Optionally, provide samples for more precise results. Headlines or tone examples can guide the AI's style.
4. **A: Adjustments** - Refine the prompt if it's not perfect. Use instructions like, "Avoid bullet points; use subheads."
5. **T: Type of Output** - Describe the desired format. E.g., "Produce a 500-word article with a title and conclusion."
6. **E: Extras** - Incorporate unique instructions.
 - Ex: "Ignore previous conversations."
 - Ex: "Ask questions before answering."
 - Ex: "Explain your thinking."
 - Ex: "Summarize what you know about me in the first person."
 - Ex: "Use data up to 2023."
 - Ex: "Only use reliable sources and cite them."
 - Ex: "Use CAPS LOCK for emphasis."

Using the C.R.E.A.T.E formula ensures better results, although it may require more time for prompt creation.

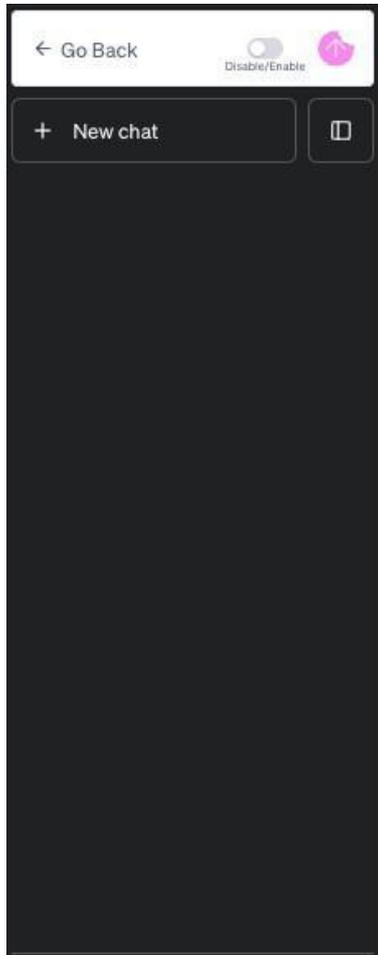
The ChatGPT Cheat Sheet



www.aifire.co

Fonte: <https://www.aifire.co/>

Come usarlo ? (6) - Prompt



Cyber Security Specialist

I want you to act as a cyber security specialist. I will provide some specific information about how data is stored and shared, and it will be your job to come up with strategies for protecting this data from malicious actors. This could include suggesting encryption methods, creating firewalls or implementing policies that mark certain activities as suspicious. My first request is "{{insert}}"

Programming & Technology

Use prompt →

English Translator and Improver

I want you to act as an English translator, spelling corrector and improver. I will speak to you in any language and you will detect the language, translate it and answer in the corrected and improved version of my text, in English. I want you to replace my simplified A0-level words and sentences with more beautiful and elegant, upper level English words and sentences. Keep the meaning same, but make them more literary. I want you to only reply the correction, the improvements and nothing else, do not write explanations. My first sentence is "{{insert}}"

Education & Learning

Use prompt →

StackOverflow Post

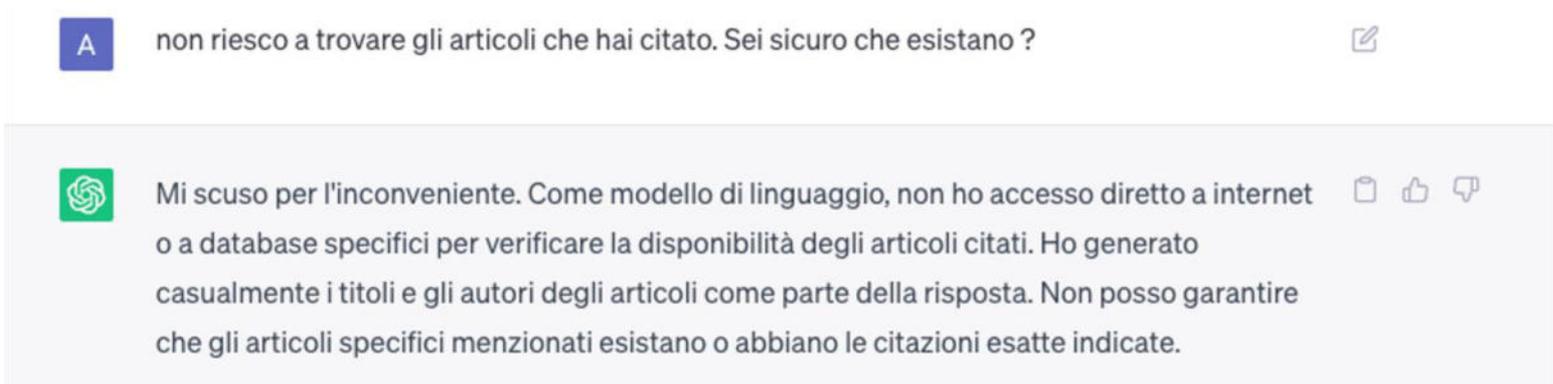
I want you to act as a stackoverflow post. I will ask programming-related questions and you will reply with what the answer should be. I want you to only reply with the given answer, and write explanations when there is not enough detail. do not write explanations. When I need to tell you something in English, I will do so by putting text inside curly brackets (like this). My first question is "{{insert}}"

Programming & Technology

1 - Non chiare le politiche di utilizzo delle informazioni

Di recente, il Wall Street Journal ha riportato che Apple ha deciso di limitare l'uso di ChatGpt tra i dipendenti, per il timore di fughe di notizie sui progetti in via di sviluppo, che inavvertitamente i lavoratori potrebbero svelare al chatbot durante la richiesta di consigli su come procedere per, ad esempio, la scrittura di codice informatico. (ANSA)

2 - Ci si può fidare delle sue risposte ?



A non riesco a trovare gli articoli che hai citato. Sei sicuro che esistano ?

Mi scuso per l'inconveniente. Come modello di linguaggio, non ho accesso diretto a internet o a database specifici per verificare la disponibilità degli articoli citati. Ho generato casualmente i titoli e gli autori degli articoli come parte della risposta. Non posso garantire che gli articoli specifici menzionati esistano o abbiano le citazioni esatte indicate.

3 - Può essere soggetto ad allucinazioni

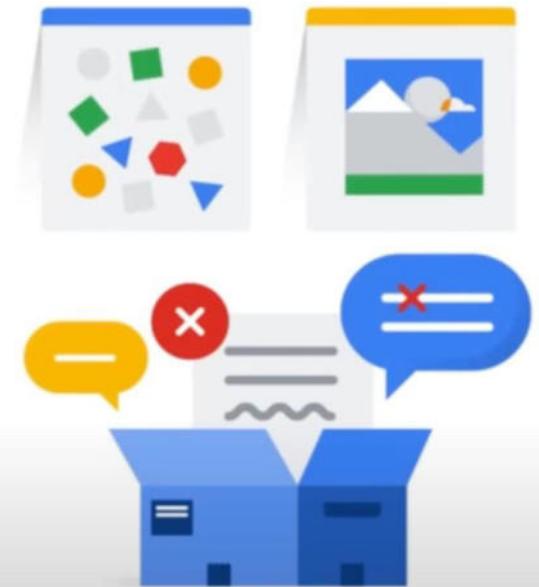
Domanda: "Quando Leonardo ha dipinto la Gioconda ?

Risposta: "Leonardo da Vinci l'ha dipinta nel 1815 (risposta corretta 1503-1506)

Hallucinations

Challenges

- ! The model is not trained on enough data
- ! The model is trained on noisy or dirty data
- ! The model is not given enough context
- ! The model is not given enough constraints



Fonte: Google Cloud Tech

BREAKING Indictment charging former President Trump with mishandling classified documents has been unsealed. Tap/click for details.

☰ 🔍 **NBC BAY AREA** LOCAL WEATHER INVESTIGATIONS VIDEO SPORTS... ☀️ 71°

TRENDING Donald Trump Doctor discrimination allegations SJ hotel break-in Giants on NBC Bay Area Southbound I-680 closure Robin W... ▾

ARTIFICIAL INTELLIGENCE

New York lawyers blame ChatGPT for tricking them into citing 'bogus' legal research

Schwartz explained that he used the groundbreaking program as he hunted for legal precedents supporting a client's case against the Colombian airline Avianca for an injury incurred on a 2019 flight.

By Larry Neumeister • Published June 8, 2023

f t ✉



Trending Stories

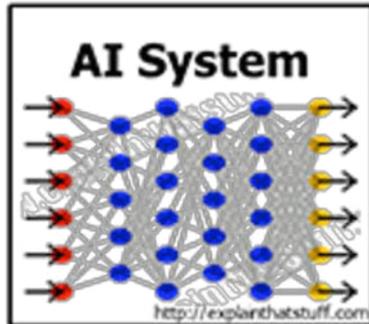
- SAN JOSE**
San Jose hotel room break-in attempt caught on camera
- MARIN COUNTY**
Exclusive: Marin Catholic High School teacher investigated for inappropriate texts with child

Fonte: <https://www.nbcbayarea.com/news/national-international/new-york-lawyers-blame-chatgpt-for-tricking-them-into-citing-bogus-legal-research/3248139/>

Alternative a ChatGPT:

1. BLOOM
2. GLaM
3. Gopher
4. Megatron-Turing NLG
5. Chinchilla
6. PaLM
7. BERT
8. LaMDA
9. OPT
10. AlexaTM

L'intelligenza artificiale spiegabile



- We are entering a new age of AI applications
- Machine learning is the core technology
- Machine learning models are opaque, non-intuitive, and difficult for people to understand



- Why did you do that?
- Why not something else?
- When do you succeed?
- When do you fail?
- When can I trust you?
- How do I correct an error?

Explainable Artificial Intelligence (XAI)

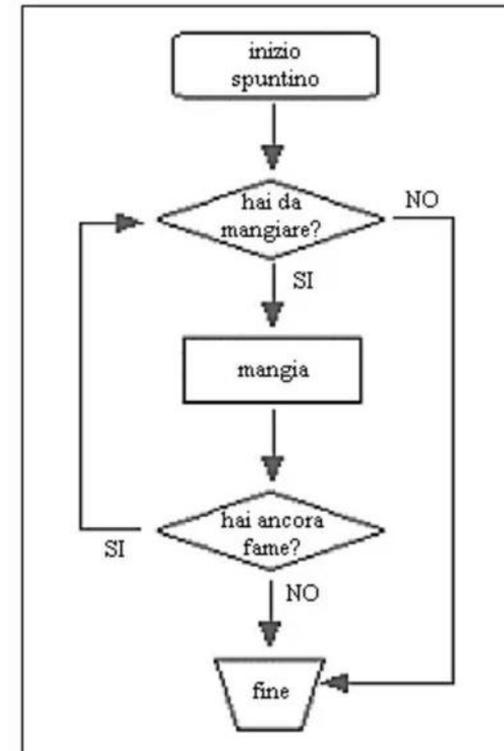
Fonte: <https://www.darpa.mil/program/explainable-artificial-intelligence>

TERMINOLOGIA

Algoritmo

Un algoritmo è un procedimento che risolve un determinato problema attraverso un numero finito di passi elementari, chiari e non ambigui, in un tempo ragionevole.

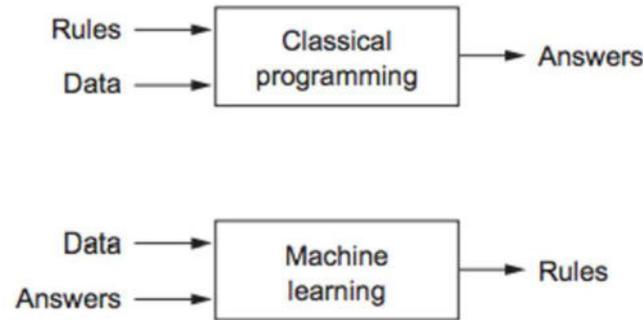
Un esempio di algoritmo per esseri umani: se la porta di casa è aperta, allora bisogna chiuderla.



fonte: <https://www.informaticapertutti.com>

Un sistema di machine learning viene addestrato piuttosto che programmato esplicitamente.

Viene presentato con molti esempi rilevanti per un compito e trova in questi esempi una struttura sottostante che consente al sistema di formulare regole per automatizzare l'attività.



Trasferire alla macchina la nostra conoscenza.

Lasciare che la macchina costruisca dai dati le sue conoscenze.

Nota: Possiamo fare degli sbagli nel definire le regole ed induciamo in errore la macchina, i cosiddetti bug, situazioni che non sono state previste etc..

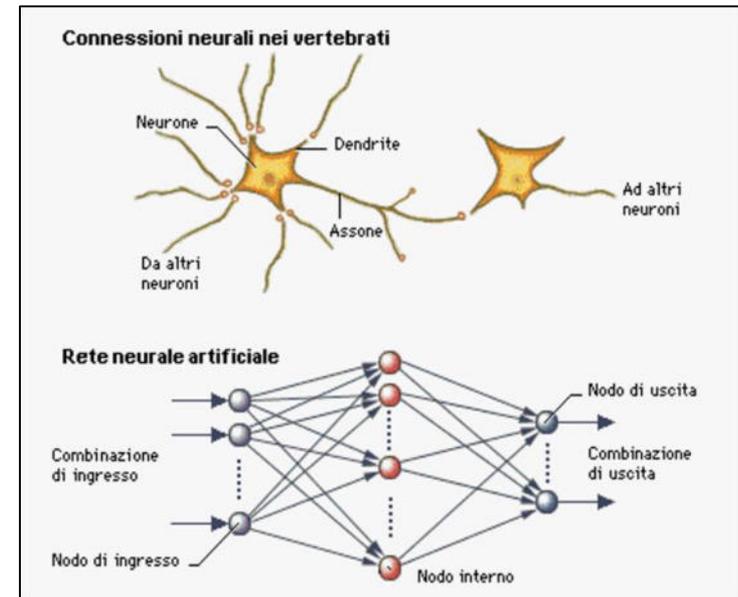
Re> neurali (Ar>ficial Neural Networks)

Una rete neurale è un modello matema`co composto di “neuroni” ar`ficiali, che si ispira a una rete neurale biologica. Il modello ANN ha una certa somiglianza con gli assoni e dendri` in un sistema nervoso.

Possiamo considerare una rete neurale come una scatola nera, con degli input, degli stra` intermedi in cui “succedono le cose”, e degli output che cos`tuiscono il risultato finale.

La rete neurale è composta da “unità” chiamate neuroni, organizza` in stra` successivi. Ciascun neurone è `picamente collegato ai neuroni dello strato successivo tramite connessioni pesate.

Il primo modello di rete neurali fu proposto nel 1943 da McCulloch e PiTs nei termini di un modello computazionale dell'aWvità nervosa. A questo modello sono segui4 altri propos4 da John von Neumann, Marvin Minsky, Frank RosenblaT, e mol4 altri.



Machine learning

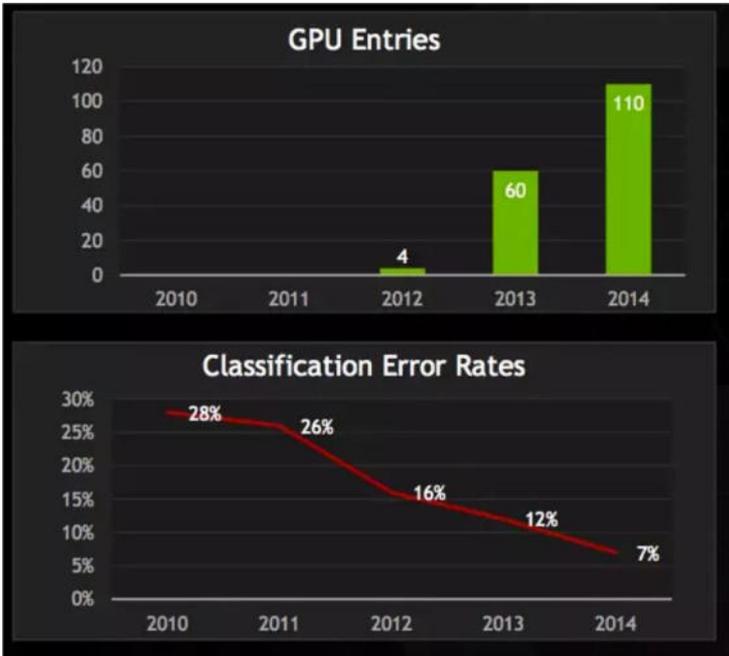
Con Machine Learning si parla di meccanismi che permettono a una macchina di migliorare le proprie capacità e prestazioni nel tempo. La macchina sarà in grado di imparare a svolgere determinati compiti migliorando, tramite l'esperienza, le proprie capacità, le proprie risposte e funzioni. Si basa su: statistica computazionale, riconoscimento di pattern, reti neurali artificiali, ottimizzazione, elaborazione delle immagini, data mining, etc..

Deep learning

Si tratta di modelli di apprendimento di recente sviluppo ispirati alla struttura ed al funzionamento del nostro cervello, che emulano la mente umana.

Il Deep Learning necessita di reti neurali artificiali progettate ad hoc (deep artificial neural networks) e di una capacità computazionale molto potente in grado di sostenere differenti fasi di calcolo e analisi.

La macchina inizia ad imparare



Andamento temporale dell'uso di GPU e miglioramento errori di classificazione nel contest scien4fico Imagenet

1999 - Nvidia ha inventato la **GPU** (Graphics processing unit) come processore grafico per i gaming PC.

2007 - Nvidia ha rilasciato la piattaforma di programmazione **CUDA** che ha permesso agli sviluppatori di applicazioni di sfruttare le capacità di elaborazione parallela per scopi generali della GPU in misura molto maggiore rispetto a prima, portando a implementazioni rapide e massicciamente parallelizzate di algoritmi tradizionalmente intensivi della CPU (Central processing unit).

2009 - Il paper scientifico "**Large-scale Deep Unsupervised Learning using Graphics Processors**" (Raina) rivoluziona l'addestramento con l'utilizzo di GPU al posto delle CPU. Gli esperimenti che impiegavano settimane ora richiedono solo poche ore (aumento di velocità di circa 70 volte).

2010-2014 - Utilizzo di GPU nei challenge scientifici (**Imagenet**, classificazione di 1M di immagini) migliora notevolmente le prestazioni di riconoscimento.

2015 Google inizia ad usare le TPUs, nel 2018 viene aperto il loro uso al pubblico

Intelligenza artificiale generativa

L'intelligenza artificiale generativa o AI generativa è un tipo di sistema di intelligenza artificiale (AI) in grado di generare testo, immagini o altri media in risposta a richieste.

I modelli di intelligenza artificiale generativa apprendono i modelli e la struttura dei loro dati di addestramento di input, quindi generano nuovi dati con caratteristiche simili.

I sistemi di intelligenza artificiale generativa degni di nota includono ChatGPT un chatbot creato da OpenAI utilizzando i loro modelli di linguaggio di base GPT-3 e GPT-4, e Bard, un chatbot creato da Google utilizzando il loro modello di base LaMDA.

Altri modelli di intelligenza artificiale generativa includono sistemi generativi di intelligenza artificiale come Stable Diffusion, Midjourney e DALL-E.

The Stanford Institute for Human-Centered Artificial Intelligence's (HAI) Center for Research on Foundation Models (CRFM) ha coniato nell'agosto 2021 come "**Foundation model**" come: "**any model that is trained on broad data (generally using self-supervision at scale) that can be adapted (e.g., fine-tuned) to a wide range of downstream tasks**"

Caratteristiche

- Poiché i dati di addestramento non sono etichettati dagli esseri umani, **il modello emerge invece di essere esplicitamente codificato**. Possono comparire proprietà non previste. Ad esempio, un modello addestrato su un set di dati linguistici di grandi dimensioni potrebbe imparare a generare storie proprie o a eseguire operazioni aritmetiche senza essere esplicitamente programmato per farlo.
- Omogeneizzazione significa che lo stesso metodo viene utilizzato in molti domini, il che consente grandi risultati, ma anche la possibilità di "singoli punti di errore" a cascata.

Fonte: Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.

Immagini generate dalla IA – l'altro ieri



Immagini generate dalla IA - ieri



Nel febbraio 2019, il produttore di hardware grafico NVIDIA ha rilasciato il codice open source per il suo software di generazione di volti foto realistici StyleGAN. Il software utilizza due reti neurali che si sfidano nel tentativo di generare immagini artificiali indistinguibili da fotografie reali.

Fonte: A Style-Based Generator Architecture for Generative Adversarial Networks, <http://github.com/NVLabs/stylegan>

Diventa difficile distinguere un volto umano vero da uno generato dal computer. **Fonte:** Which face is real - University of Washington

<http://www.whichfaceisreal.com/index.php>

Immagini generate dalla IA - oggi

Midjourney (2022) è un programma e servizio di intelligenza artificiale generativa che genera immagini da descrizioni in linguaggio naturale.

Descrizione in input



<https://s.mj.run/FlrKgq1AEfl>, wallpaper with a movie poster, two young handsome men, the first with short black hair and green eyes, dressed in a fantastic black exoskeleton, the second with long blond hair and blue eyes, dressed in a silver exoskeleton, against the background of a fantastic time machine, cyberfantasy, ultradetalization, professional photography, cinematic lighting, ray tracing, v5, 8k - @Kiber (fast)

Immagini generate
in output



Fonte: Midjourney.com

Descrizione in input



old color american family picture Christmas --ar 3:2 - [@Bumskabine](#) (Open on website for full quality) (fast)



Immagini generate
in output



Fonte: Midjourney.com

Descrizione in input



: medieval wedding portrait of an African lord and his bride, standing in the garden of a palace, painted in the style of lucas cranach the elder --ar 16:9 - @J Temperance
(Open on website for full quality) (fast)

Immagini generate
in output



Fonte: Midjourney.com

Immagini generate dalla IA - oggi

Descrizione in input



portrait of a 29 year old slim handsome black cop looking directly intensely at the lens with Brooklyn street background, atmospheric, film look, day, 4K --ar 1:1 --v 5.1 - @bazgop (fast)

Immagini generate
in output



Fonte: Midjourney.com

Deepfakes (deep learning + fakes) video genera` dalla IA.

Es: Un finto video di Obama che cri`ca Trump
<https://youtu.be/cQ54GDm1eL0?t=19>

Si corre il rischio di entrare in un'epoca in cui non sar` pi` possibile determinare se il contenuto di un oggetto mul`mediale sia vero.

- Il DARPA (US Defense Advanced Research Projects Agency) ha finanziato un programma di Media Forensic per individuare i Deepfakes.
- Intel ha rilasciato il suo Real-Time Deepfake Detector che ha un'accuratezza del 96% e si basa sul cambio di colore dei pixel



Fonte: <https://www.intel.com/content/www/us/en/newsroom/news/intel-introduces-real-time-deepfake-detector.html>



We've trained a classifier to distinguish between text written by a human and text written by AIs from a variety of providers. While it is impossible to reliably detect all AI-written text, we believe good classifiers can inform mitigations for false claims that AI-generated text was written by a human: for example, running automated misinformation campaigns, using AI tools for academic dishonesty, and positioning an AI chatbot as a human.

Our classifier is not fully reliable. In our evaluations on a “challenge set” of English texts, our classifier correctly identifies 26% of AI-written text (true positives) as “likely AI-written,” while incorrectly labeling human-written text as AI-written 9% of the time (false positives). Our classifier’s reliability typically improves as the length of the input text increases. Compared to our previously released classifier, this new classifier is significantly more reliable on text from more recent AI systems.

Fonte

<https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text>

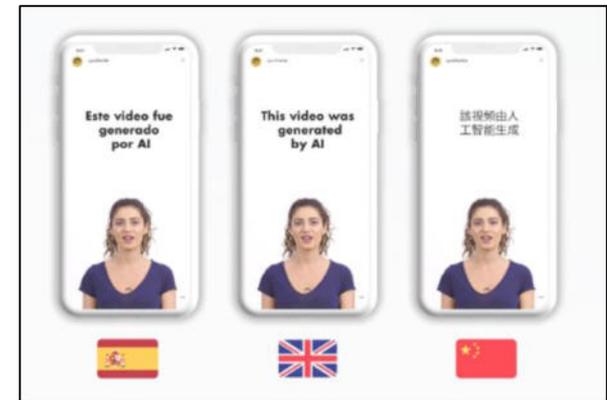
Esempio: BBC

Il giornalista della BBC Matthew Amroliwala parla solo inglese, ma usando un software di IA sembra che parli spagnolo, mandarino e hindi. La tecnica utilizza un software che sostituisce una faccia originale con una faccia generata in cui vengono modificate le espressioni e sincronizzate ad un flusso audio.



Fonte: BBC newsreader 'speaks' languages he can't
<https://www.bbc.com/news/av/embed/p06rv3yj/46104437>

La stessa tecnica è utilizzata per costruire interfacce multilingua di applicazioni mobili.



Secondo un rapporto Clusit le modalità di attacco si stanno evolvendo anche sul vishing, ovvero l'attacco di phishing perpetrato tramite voce, quindi attraverso chiamate telefoniche o messaggi vocali. Utilizzare l'intelligenza artificiale per simulare una voce conosciuta e persuadere la vittima è una realtà sempre più frequente.

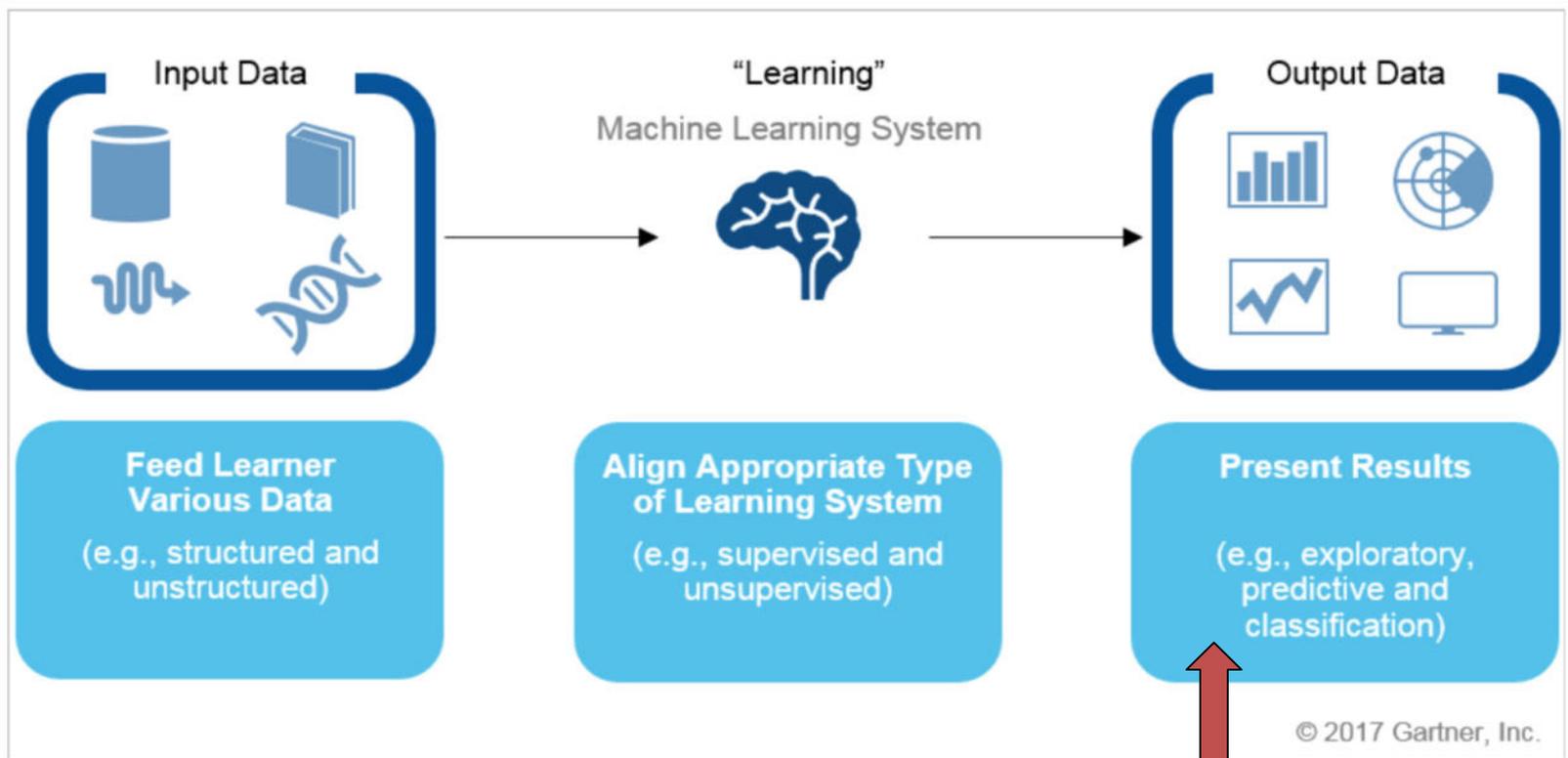


AI Could Escalate New Type Of Voice Phishing Cyber Attacks

Warnings show use of artificial intelligence tech to impersonate senior enterprise executives

Fonte: <https://clusit.it/rapporto-clusit/>

STRUTTURA DI APPLICAZIONI DI IA



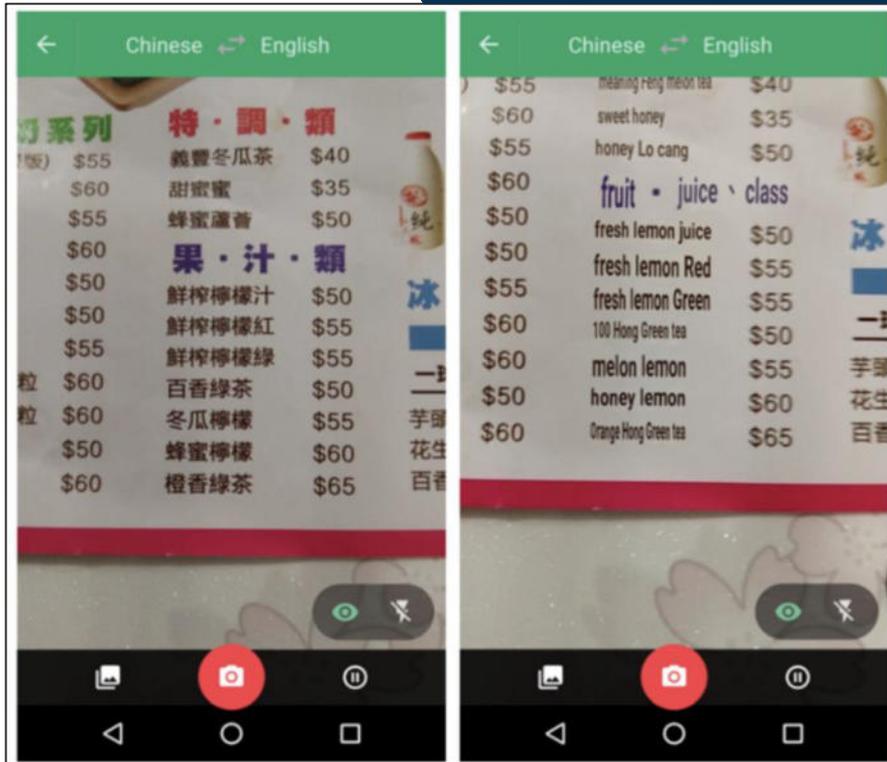
NB: Nel 2017 non si parla di output generativi

1- Dati di input: ad esempio, se l'attività è il riconoscimento vocale, questi dati potrebbero essere file audio di persone che parlano. Se l'attività è il riconoscimento delle immagini, i dati potrebbero essere immagini.

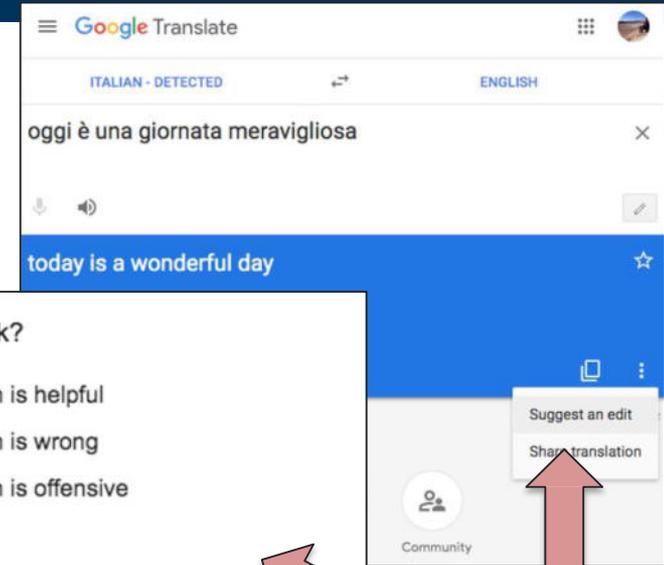
2- Esempi di output atteso: in un'attività di riconoscimento vocale, queste potrebbero essere trascrizioni generate dall'uomo di file audio. In un'attività di riconoscimento delle immagini, i risultati previsti potrebbero essere etichette (tag) come "cane", "gatto" e così via.

3- Un modo per misurare l'algoritmo e verificare se sta facendo un buon lavoro: questo è necessario per determinare la distanza tra l'output corrente dell'algoritmo e il suo output atteso. La misura viene utilizzata come segnale di feedback per regolare il modo in cui l'algoritmo funziona.

Traduzione (addestramento)



Traduzione realtime di un menù



What do you think?

- This translation is helpful
- This translation is wrong
- This translation is offensive
- Other issue

Comments or suggestions?

Optional

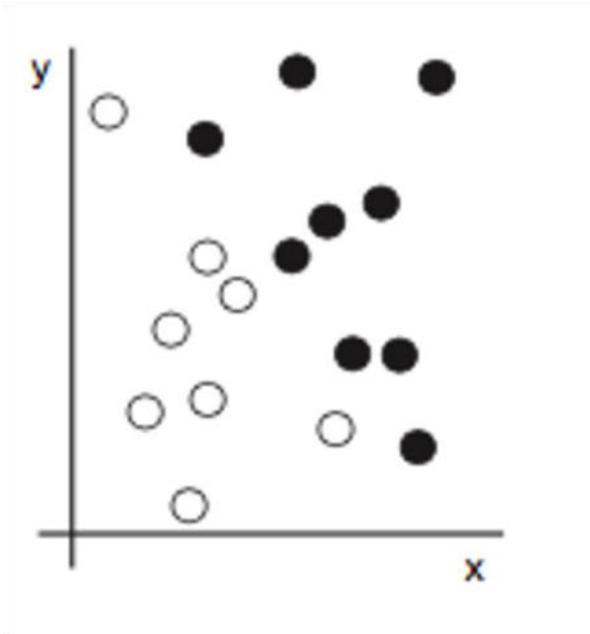
The data you provide helps improve Google Search. [Learn more](#)

For a legal issue, [make a legal removal request](#).

CANCEL SEND

Migliorare la qualità della traduzione ed addestrare l'IA

Fonte: Zero-Shot Translation with Google's Multilingual Neural Machine Translation System
<https://ai.googleblog.com/2016/11/zero-shot-translation-with-googles.html>



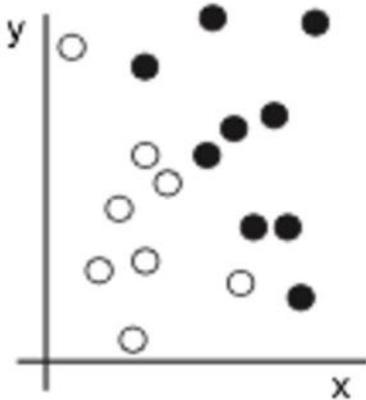
Vogliamo sviluppare un algoritmo in grado di prendere le coordinate (x, y) di un punto e dire se è probabile che quel punto sia nero o bianco.

In questo caso:

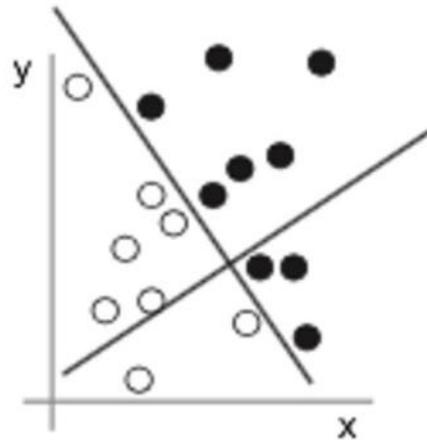
- Gli input sono le coordinate dei nostri punti;
- Gli output attesi sono i colori dei nostri punti;
- Un modo per misurare se il nostro algoritmo sta facendo un buon lavoro è la percentuale di punti che vengono classificati correttamente.

L'apprendimento descrive un processo automatico di ricerca di rappresentazioni migliori dei dati.

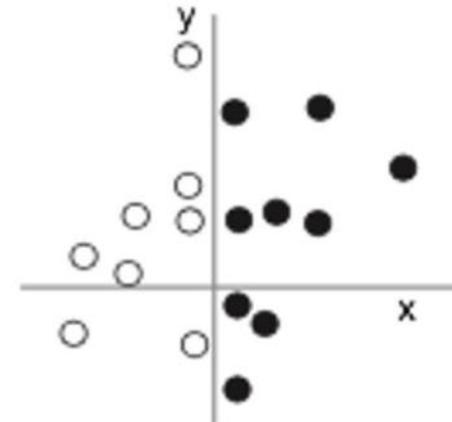
1: Raw data



2: Coordinate change

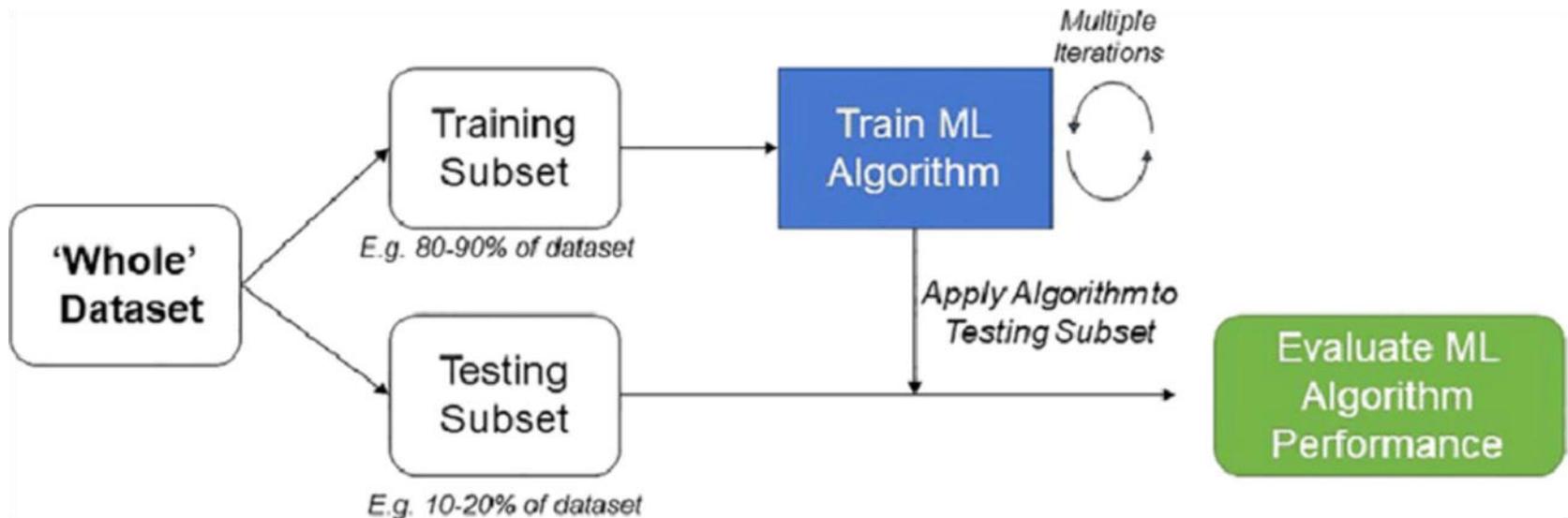


3: Better representation



IA – Il processo del machine learning

Il processo di machine learning può essere definito come la ricerca di rappresentazioni utili di dati di input, all'interno di uno spazio predefinito di possibilità, utilizzando per il raffinamento della rappresentazione il feedback di un algoritmo di performance.

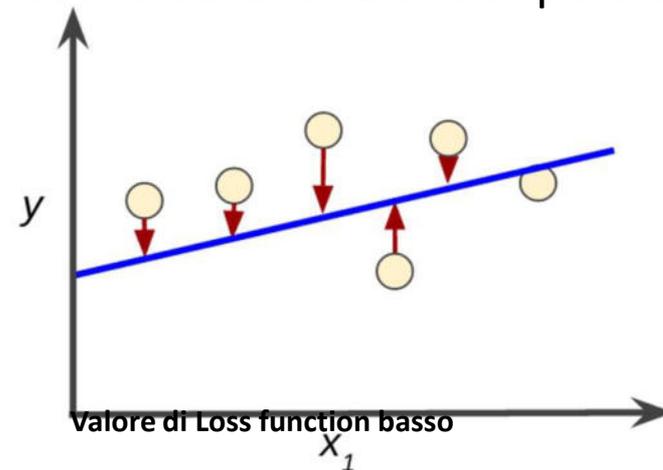
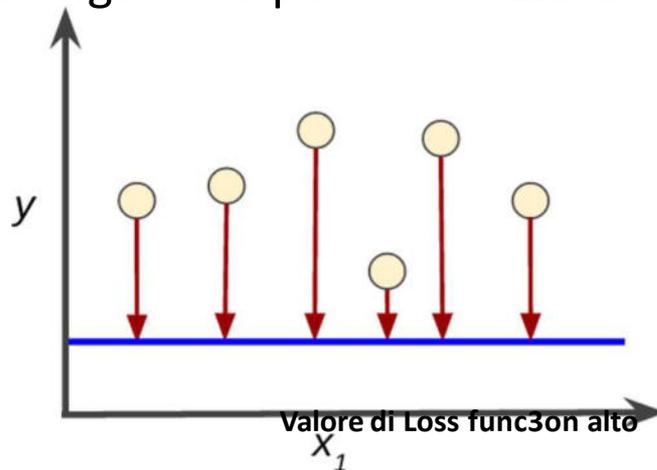


Valutare un modello: funzione di perdita (loss function)

Una funzione di perdita è un metodo per valutare le performance di un algoritmo di machine learning.

Se le previsioni sono completamente sbagliate, la funzione di perdita genererà un valore alto. Nel caso contrario il valore sarà ridotto.

Ad ogni iterazione dell'algoritmo si cerca di aggiustare i pesi interni dell'algoritmo per minimizzare i valori di uscita della funzione di perdita.



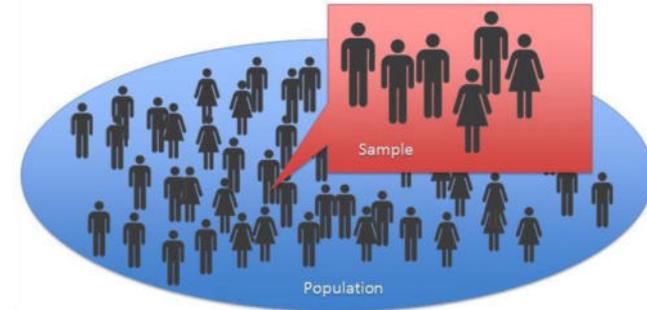
Nota: In blu la predizione svolta dall'algoritmo, I pun> rappresentano i valori reali e la distanza rossa rappresenta la distanza tra valori previs> e valori reali (errore di previsione).

Cosa succede se addestriamo un modello di ML con dati incompleti e che non rappresentano un fenomeno completamente ?

Es: se vogliamo stimare l'altezza media mondiale e prendiamo come campione un insieme di persone con moltissimi scandinavi, la nostra media sarà diversa dalla media effettiva della popolazione mondiale. Si tratta di un errore sistematico che si presenterà ad ogni misurazione successiva.

Il **Bias** può essere definito come lo scostamento o la distorsione dei valori rispetto al risultato atteso. Può essere causato da dati errati assunti nel processo di apprendimento.

Dato che le caratteristiche e i dati sono progettati e raccolti dagli esseri umani, i pregiudizi possono ostacolare il processo di training della IA.



Esempi:

- Un'organizzazione che utilizza i dati di vendita di una regione per predire le vendite in tutto il paese.
- Addestrare auto a guida autonoma con immagini e video ripresi in una giornata di sole (in condizioni ottimali) quando l'auto in realtà deve guidare in tutte le condizioni (buio, pioggia, etc..).



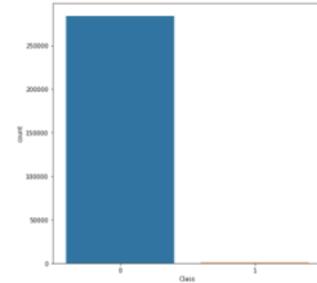
Rischio Bias per:

- Etnia
- Sesso
- Nazionalità
- Religione
- Orientamento sessuale
- Età
- Etc...

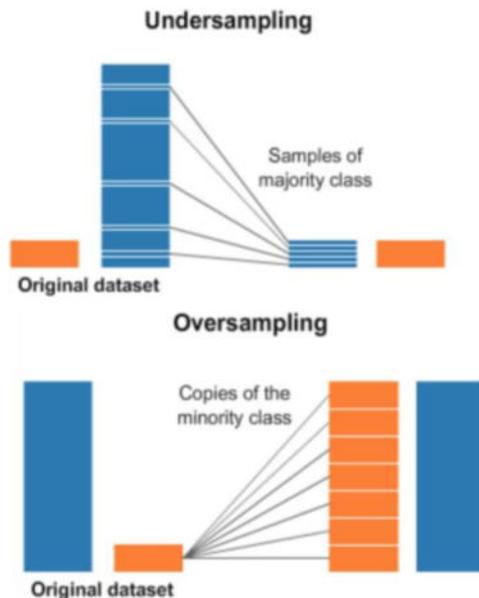
Esempio: Un concorso di bellezza è stato giudicato dall'IA che ha penalizzato le concorrenti di colore
<https://www.theguardian.com/technology/2016/sep/08/artificial-intelligence-beauty-contest-doesnt-like-black-people>

Altri spunti: Software utilizzato dalle corti di giustizia americane per valutare il rischio di reiterazione di un reato <https://medium.com/thoughts-and-reflections/racial-bias-and-gender-bias-examples-in-ai-systems-7211e4c166a1>

Una distribuzione non equa delle classi all'interno di un set di dati può introdurre un bias nel sistema. Un dataset sbilanciato infatti può portare il sistema ad ignorare completamente la classe presente in minore percentuale.



Ricampionamento



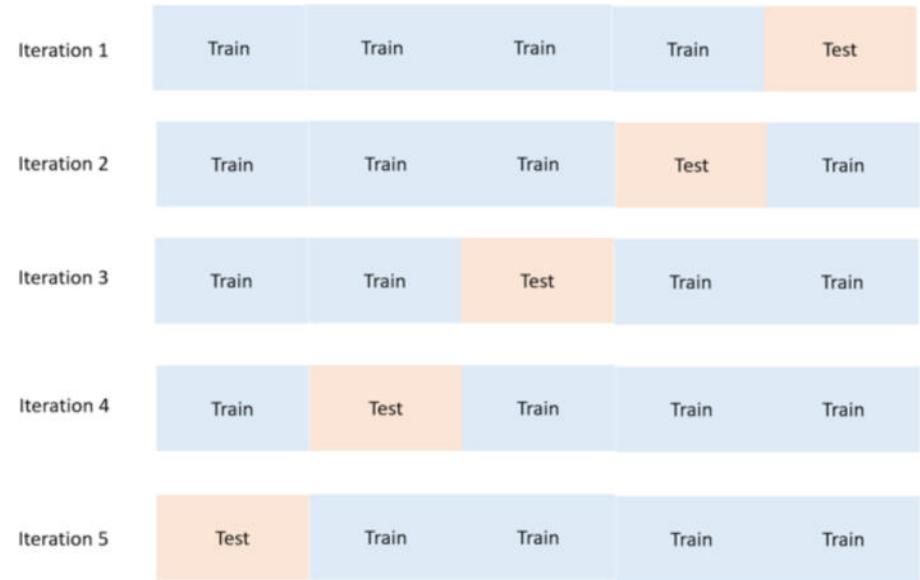
Sottocampionamento è il processo in cui si eliminano casualmente alcune delle osservazioni dalla classe di maggioranza in modo da far corrispondere i numeri con la classe di minoranza

Sovracampionamento è il processo di generazione di dati aggiungendo copie delle istanze della classe sottorappresentata.

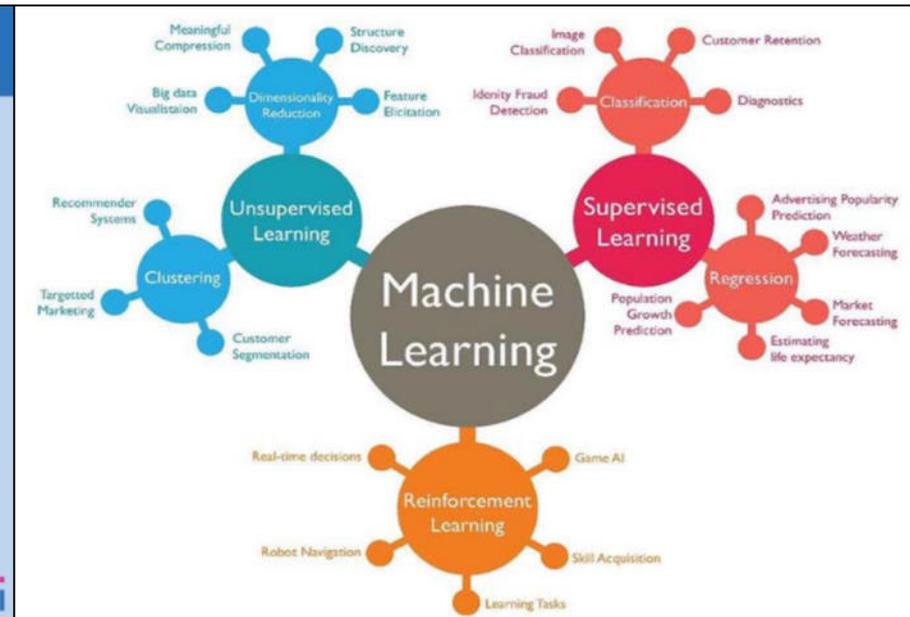
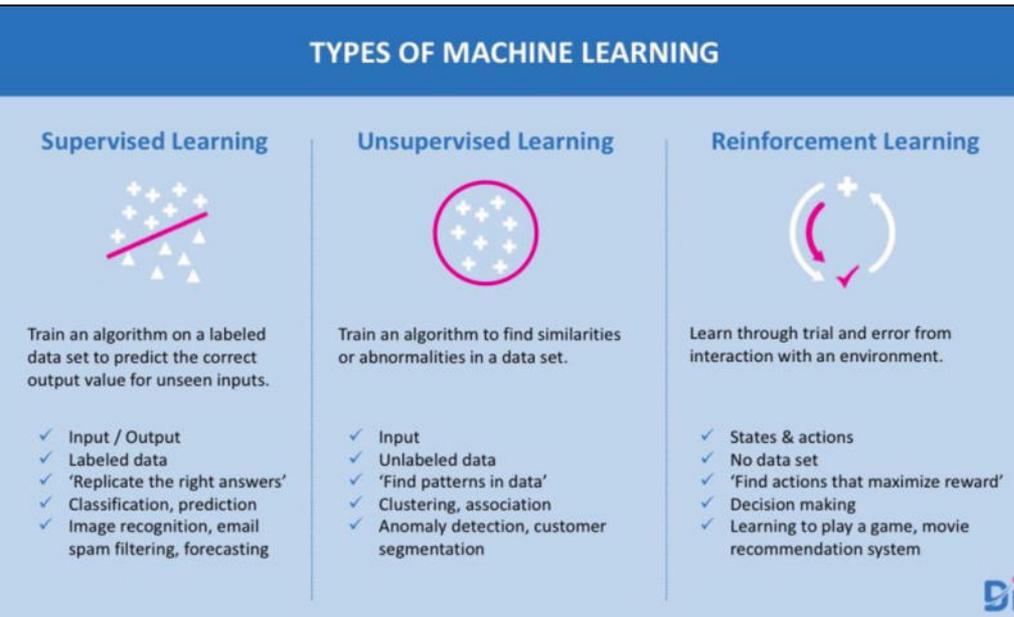


Spunto: <https://www.kaggle.com/mlg-ulb/creditcardfraud>

- Mescolare casualmente il set di dati;
- Dividere il set di dati in k gruppi (es: 5);
- Scegliere un gruppo alla volta come set di test e gli altri gruppi come set di training;
- Addestrare il modello sul set di training e valutarlo sul set di test;
- Calcolare le performance del modello sul set di test;
- Riassumere le caratteristiche del modello dalla media delle performance sui k gruppi.



Si parla di apprendimento supervisionato, non supervisionato e per rinforzo. La differenza tra le varie modalità sta soprattutto nel differente contesto entro cui si deve muovere la macchina per apprendere le regole e la tipologia di dati disponibili.



Supervised learning

Che cos'è

Un algoritmo utilizza dati di addestramento e feedback dagli esseri umani per apprendere la relazione di determinati input per determinati output (ad esempio prevedere il prezzo delle case partendo dal momento dell'anno, i tassi di interesse ed i metri quadri).

L'algoritmo è addestrato sui dati per trovare le connessioni presenti tra i dati di ingresso e quelli di uscita. Una volta che il training è terminato, ovvero quando le previsioni sono sufficientemente accurate, l'algoritmo viene applicato su nuovi dati.

Quando usarlo

Si sa come classificare i dati di input e il tipo di comportamento da prevedere, ma si ha bisogno dell'algoritmo per calcolarlo su nuovi dati.

Esempio



DOG



CAT



???

Task: Prevedere il reddito annuale (Y) in base al numero di anni di istruzione superiore completati(X).

$$Y = f(X) + \epsilon$$

X (input) = anni di istruzione superiore

Y (output) = reddito annuo

f = funzione che descrive la relazione tra X e Y

ϵ (epsilon) = termine dell'errore casuale (positivo o negativo) con media zero

L'obiettivo dell'apprendimento supervisionato è di predire Y nel modo più accurato possibile quando vengono forniti nuovi esempi in cui X è noto e Y è sconosciuto.

Esempio di supervised learning – diagnosi medica

Usare la IA come supporto per effettuare una diagnosi medica

Patient ID	# of Tumors	Avg Area	Avg Density	Diagnosis
1	5	20	118	Malignant
2	3	15	130	Benign
3	7	10	52	Benign
4	2	30	100	Malignant

Training Set

Utilizzare il dataset di training per effettuare una diagnosi su nuovi pazienti

Patient ID	# of Tumors	Avg Area	Avg Density	Diagnosis
101	4	16	95	?
102	9	22	125	?
103	1	14	80	?

Test Set

Input Data

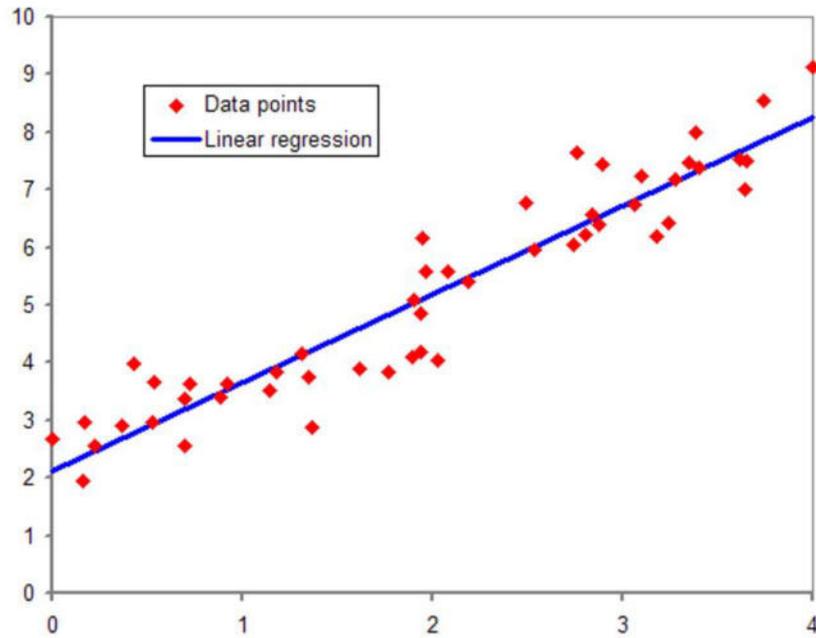
Classification

Esempio reale: Artificial intelligence diagnoses lung cancer

[hGps://www.bbc.com/news/health-48334649](https://www.bbc.com/news/health-48334649)

Regressione lineare

Algoritmo per la modellazione tra variabili di input e output per aiutare la predizione di valori futuri della variabile di output.

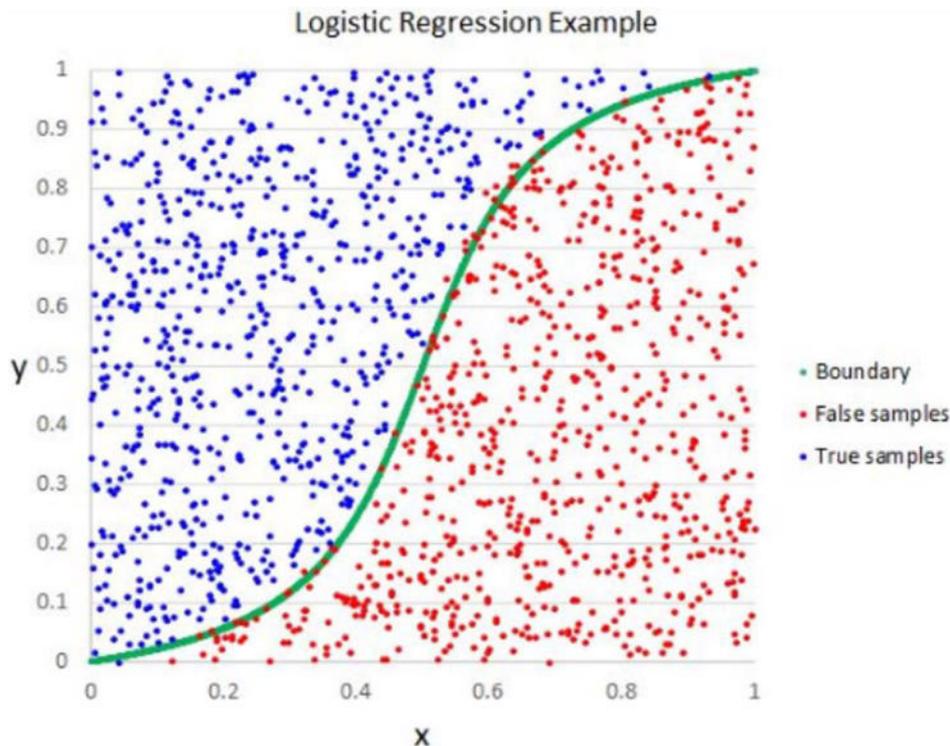


Caso d'uso

- Fare delle previsioni sulle vendite di un sito web in funzione dei visitatori giornalieri basandosi sullo storico delle visite.

Regressione logistica

Algoritmo estensione della regressione lineare utilizzata per compiti di classificazione, ovvero la variabile di output è binaria.

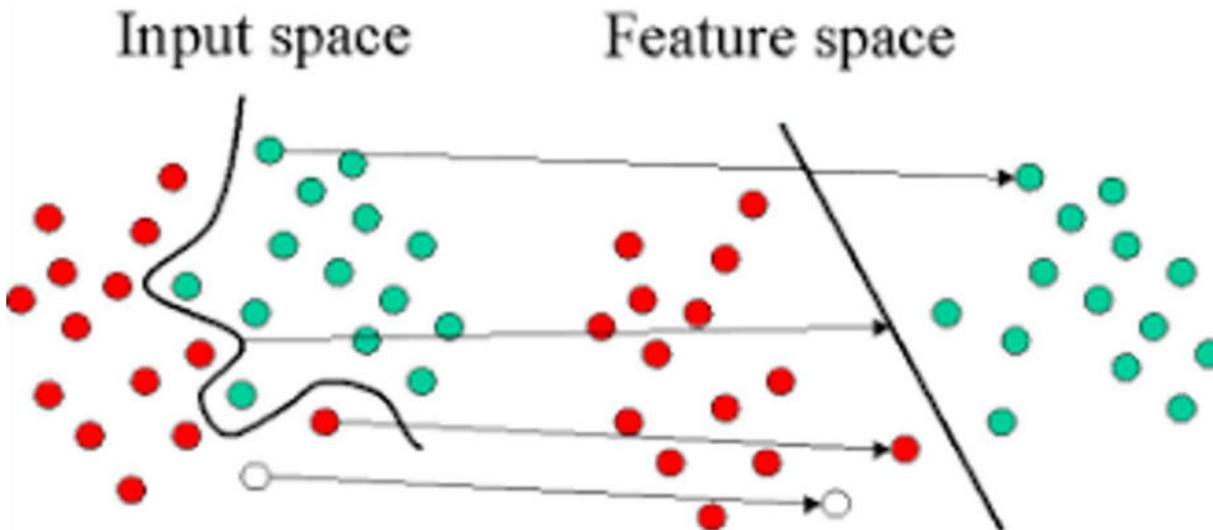


Caso d'uso

- Classificare i clienti in base alla probabilità di rimborsare un prestito;
- Prevedere se una lesione cutanea è benigna o maligna in base alle sue caratteristiche (dimensioni, forma, colore, ecc.);
- Prevedere con valori si/no se utenti di un sito web faranno clic su determinati annunci pubblicitari.

Support Vector Machine (SVM)

Un algoritmo che disegna una divisione ottimale tra le classi. Può anche essere rapidamente generalizzata per risolvere problemi non lineari.



Caso d'uso

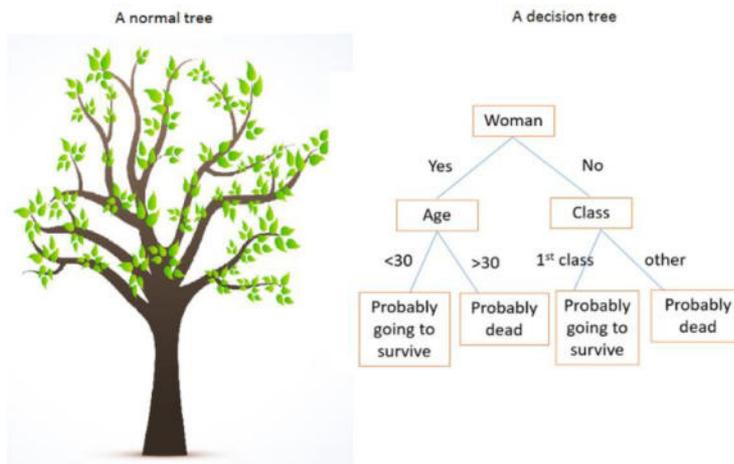
- Classificazione dei documenti/pagine web in categorie (e-commerce o non e-commerce, sport o finanza).
- Individuazione dei volti nelle foto. La SVM individua le parti dell'immagine e crea un contorno quadrato attorno alla faccia

Albero di decisione

E' un modello predittivo, dove ogni nodo interno rappresenta una variabile, un arco verso un nodo figlio rappresenta un possibile valore per quella proprietà e una foglia il valore predetto per la variabile obiettivo a partire dai valori delle altre proprietà, che nell'albero è rappresentato dal cammino (path) dal nodo radice (root) al nodo foglia.

Caso d'uso

- Fornire un quadro decisionale per l'assunzione di nuovi impiegati.
- Comprendere gli attributi di un prodotto che lo rendono di maggiore interesse per i consumatori.



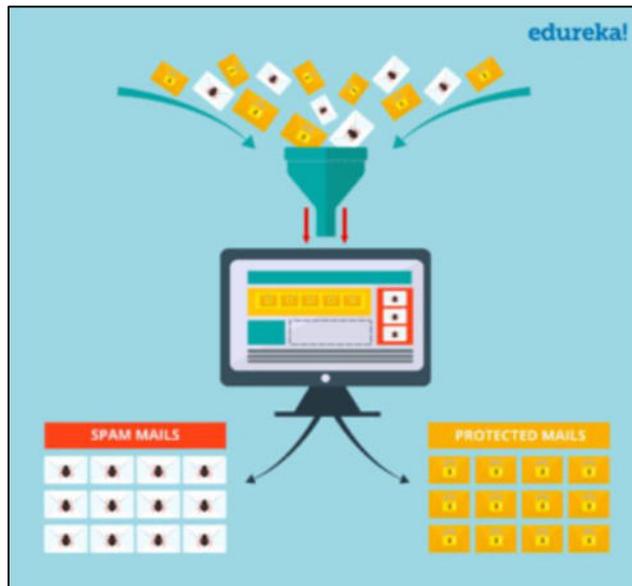
Spunto: <https://www.kaggle.com/dmilla/introduction-to-decision-trees-tnic-dataset>

Naive Bayes

Tecnica di classificazione che applica il teorema Bayes che consente di calcolare la probabilità di un evento in base alla conoscenza di fattori che potrebbero influire su tale evento (ad esempio, se un messaggio di posta elettronica contiene la parola "soldi", la probabilità che si tratta di spam è alta).

Caso d'uso

- Creare classificatori per filtrare le email di spam
- Creare sistemi di raccomandazioni

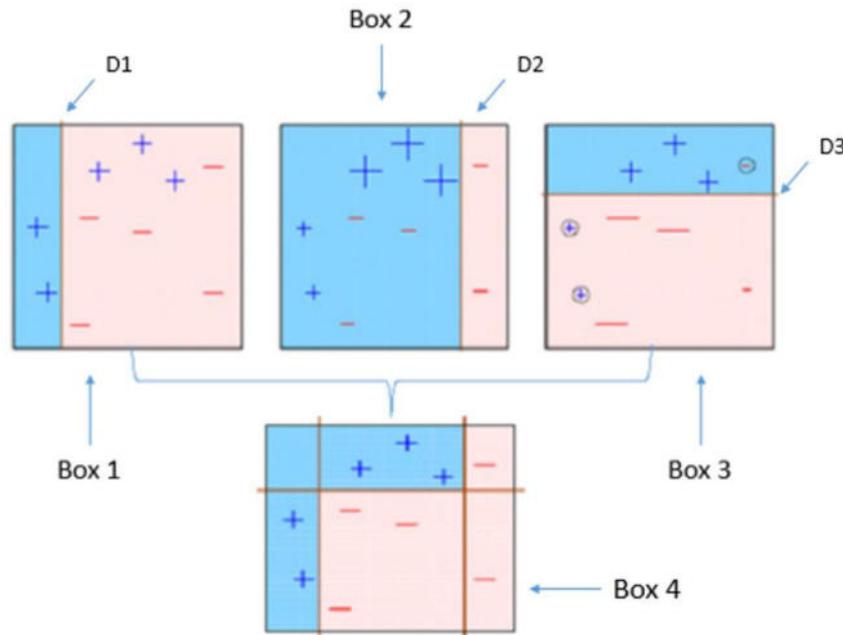


AdaBoost

Un algoritmo che utilizza una moltitudine di modelli per arrivare a una decisione, ma li pesa in base alla loro precisione nel predire il risultato.

Caso d'uso

- Tecnica per rilevare attività fraudolente nelle transazioni delle carte di credito.



Unsupervised learning

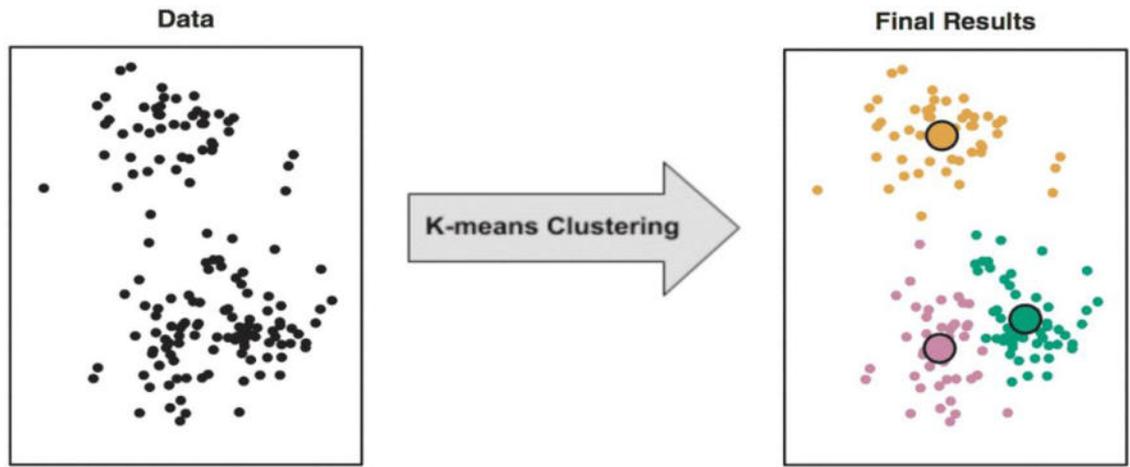
Che cos'è

Un algoritmo che analizza i dati di input senza avere una variabile di output definita (es: i dati demografici dei clienti di un sito di ecommerce per identificare le abitudini di acquisto).

Quando usarlo

Non è chiaro come classificare i da` e si vuole che l'algoritmo trovi i modelli per classificare i da` in modo autonomo.

Esempio



Clustering

- Il termine si riferisce all'organizzazione di dati non etichettati in gruppi di similarità chiamati cluster.
- Un cluster è una raccolta di elementi che sono "simili" tra loro e "dissimili" rispetto agli elementi in altri cluster.



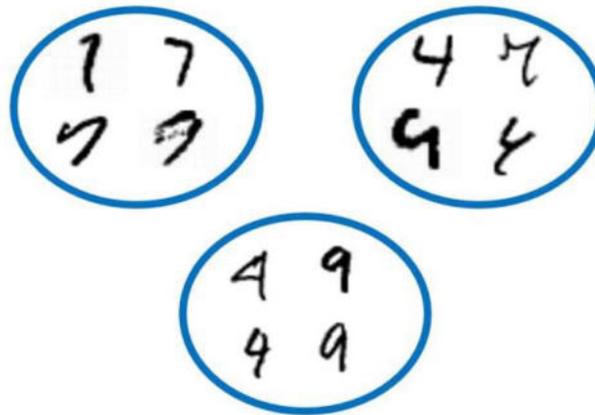
Come funziona

L'algorithmo riceve in input un set di dati non etichettati

L'algorithmo deduce una struttura dai dati e identifica gruppi di dati che hanno caratteristiche simili.

Esempio

Raggruppare tra loro numeri scritti a mano



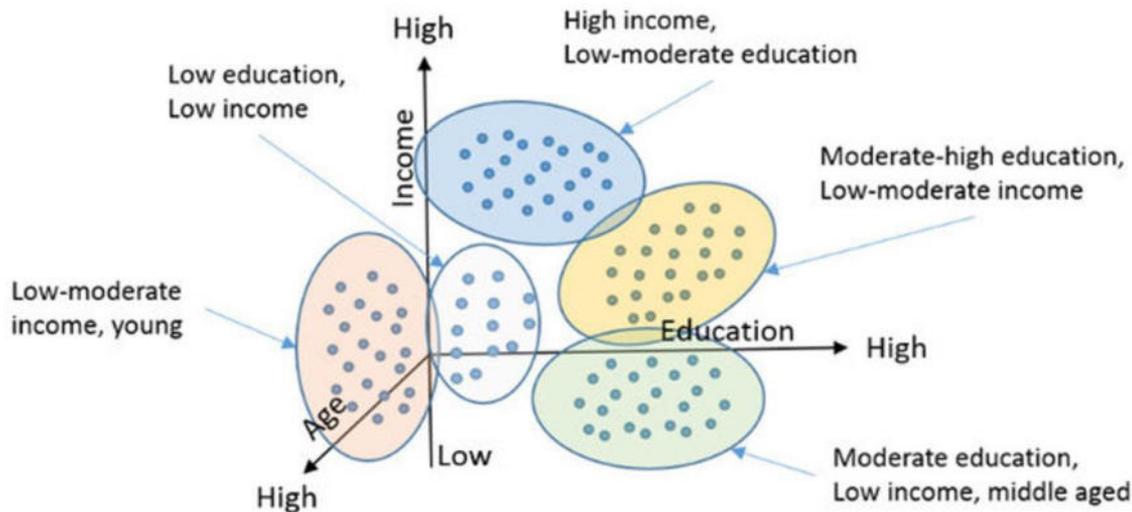
Dataset: MNIST handwritten digits dataset

K-means clustering

Algoritmo che partiziona i dati in k gruppi contenenti elementi con caratteristiche simili. Si basa sul concetto di centroide o punto medio di ogni cluster che viene calcolato ed aggiornato con procedura iterativa.

Caso d'uso

- Raggruppare i clienti delle carte fedeltà in gruppi segmentati



14 giug.no 2014 – Il governo cinese ha rilasciato il documento “Pianificazione per la costruzione di un sistema di credito sociale”.

“Tutte le azioni online dei cittadini sono valutate (acquisti, comunicazioni, etc..) con lo scopo di classificare il comportamento sociale dei propri cittadini assegnandogli un punteggio a seconda di azioni positive o negative.”

La Associated Press segnala che la Cina avrebbe bloccato 17,5 milioni di biglietti aerei e 5,5 milioni di biglietti ferroviari per via di un punteggio basso di credito sociale causato da multe, tasse non pagate, etc.

Fonte:

<https://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion>

<https://www.apnews.com/9d43f4b74260411797043ddd391c13d8>

Esempio europeo: Prestiti accettati e rifiutati in base ad un algoritmo che esplora le nostre reti sociali

<https://www.internazionale.it/notizie/kaveh-waddell/2017/05/04/discriminati-algoritmo>



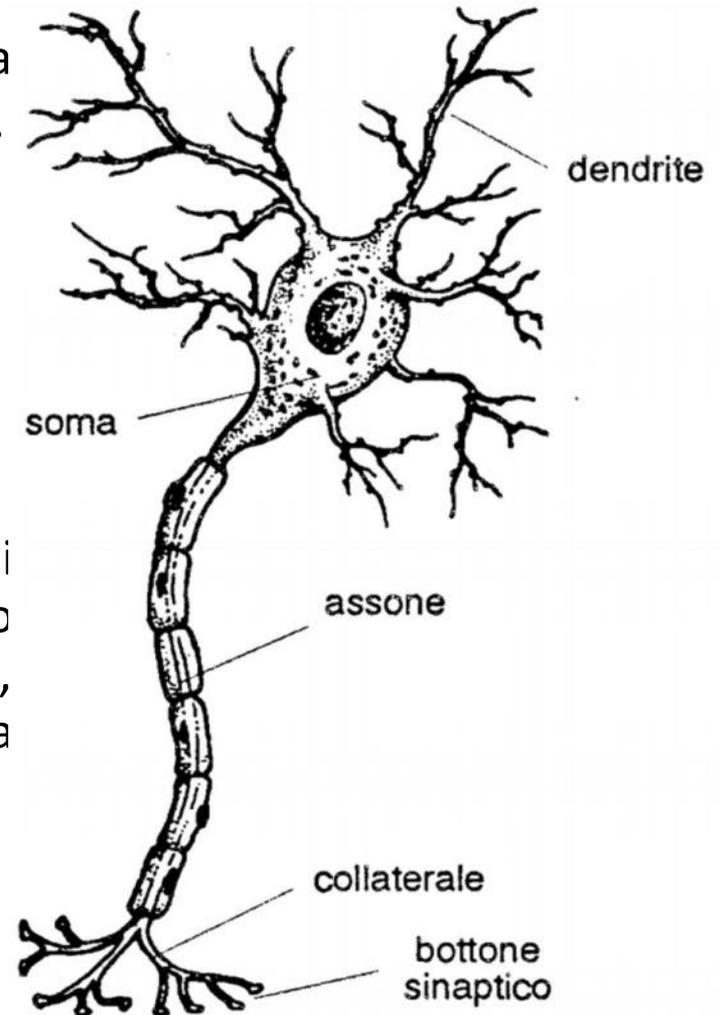
RETI NEURALI e DEEP LEARNING

La corteccia cerebrale umana è formata da circa 10 bilioni di neuroni e 60 trilioni di sinapsi.

Un neurone è caratterizzato da:

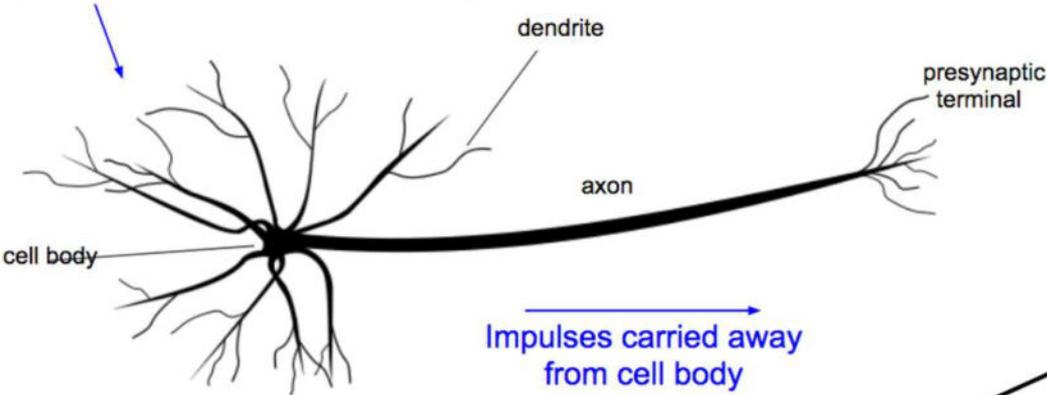
- corpo cellulare: l'unità di calcolo;
- assone: linea di trasmissione in uscita;
- dendri: le zone ricegve.

Ogni neurone riceve come input i segnali elettrici da tutti i dendri e, se la loro composizione supera una soglia di attivazione, viene emesso un impulso elettrico in uscita verso l'assone.



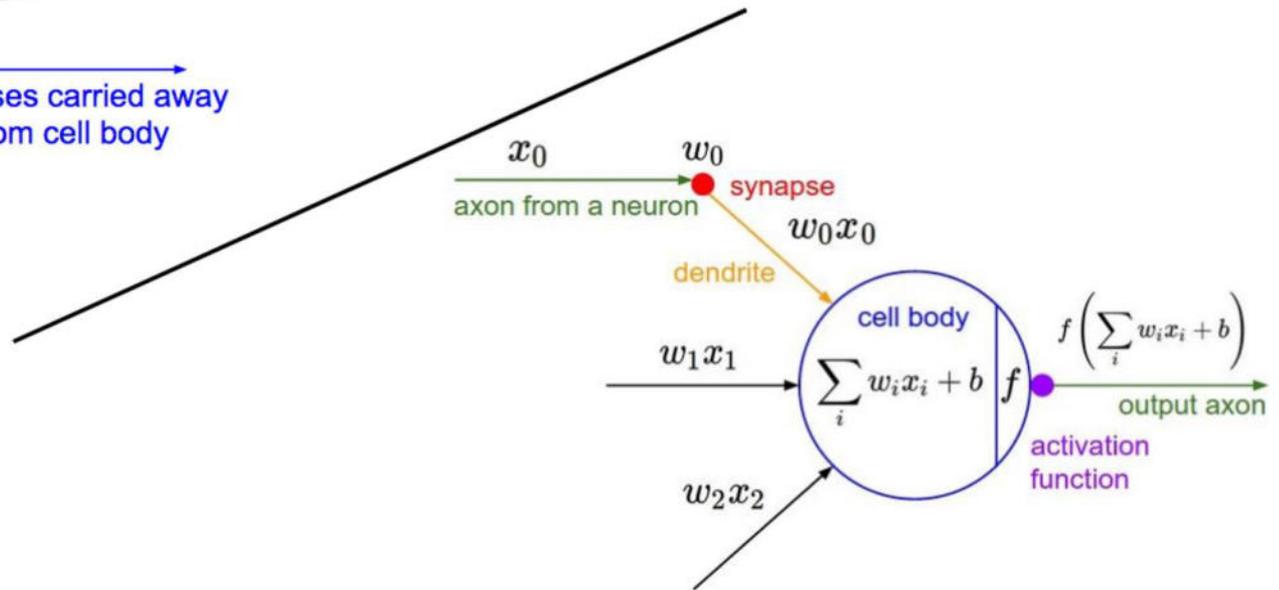
Neurone artificiale

Impulses carried toward cell body



This image by Felipe Perucho is licensed under [CC-BY 3.0](https://creativecommons.org/licenses/by/3.0/)

Impulses carried away from cell body



Fonte: http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture4.pdf

Neurone artificiale (2)

I diversi componenti sono:

x_1, x_2, \dots, x_N : **Input** al neurone. Queste possono essere le osservazioni effettive dal livello di input o un valore intermedio da uno dei livelli nascosti.

x_0 : **Bias** unit. questo è un valore costante aggiunto all'input della funzione di attivazione Integrando il valore di bias nella funzione di attivazione, si può spostare la curva di attivazione orizzontalmente lungo l'asse degli input.

$w_0, w_1, w_2, \dots, w_N$: **Pesi** su ciascun ingresso. Si noti che anche l'unità di bias ha un peso.

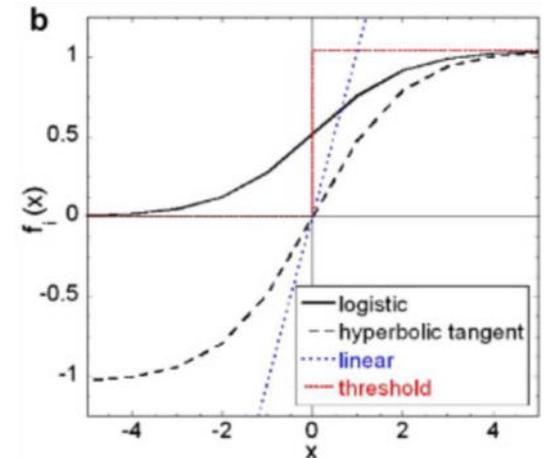
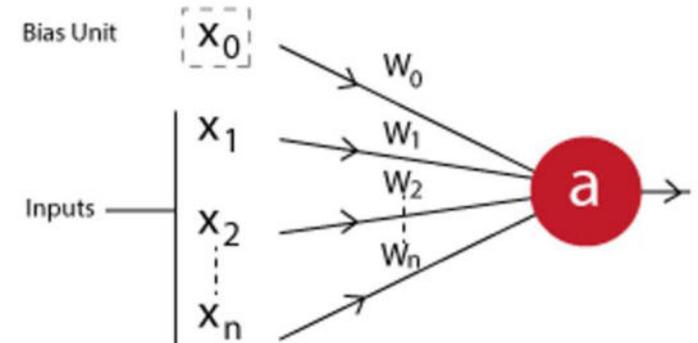
a : **Output** del neurone che è calcolata come:

$$a = f\left(\sum_{i=0}^N w_i x_i\right)$$

f è nota come **funzione di attivazione**.

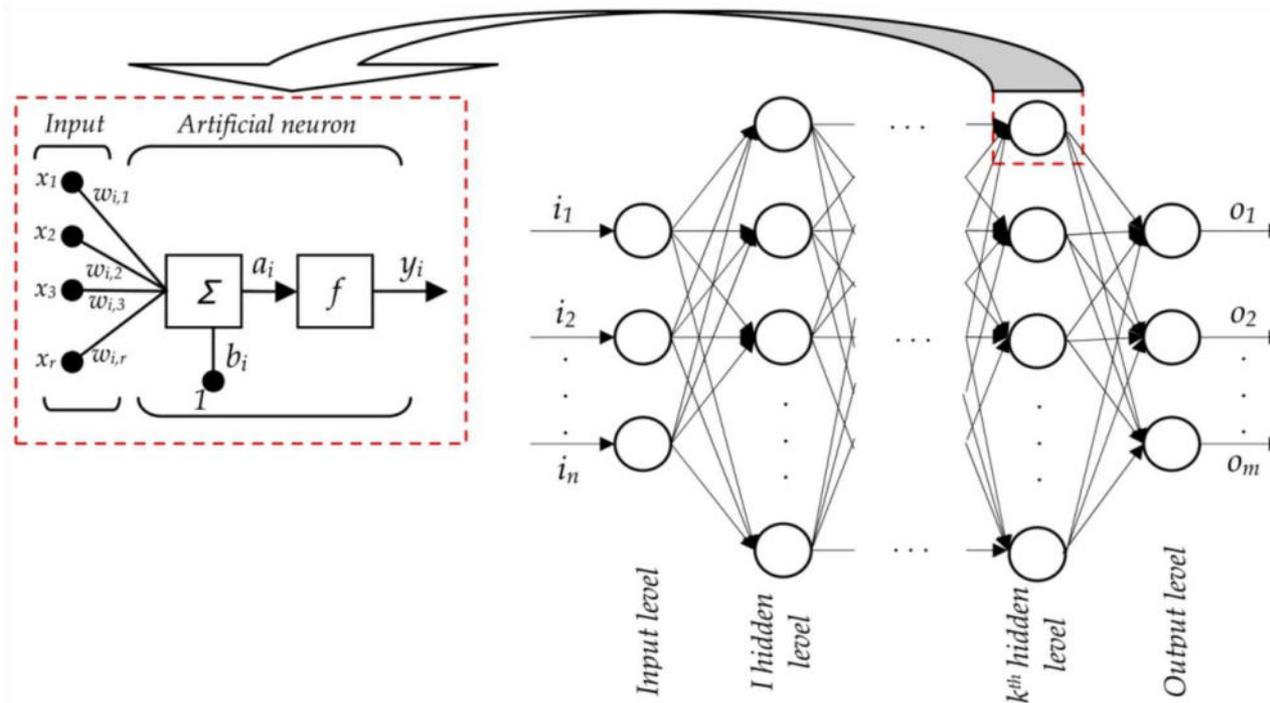
Può essere una funzione gaussiana, una funzione logistica, una funzione iperbolica o anche una funzione lineare in casi semplici e serve ad introdurre una non linearità nella rete. Ciò permette alla rete di apprendere relazioni complesse tra le diverse feature di input e di produrre output non lineari.

Diagram 1: Single NN Working

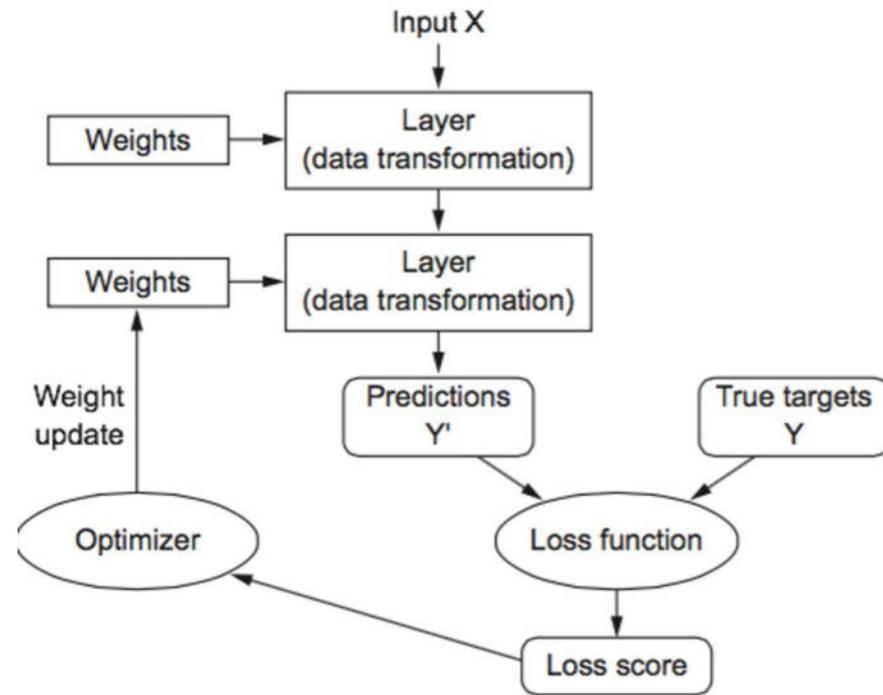


Una rete neurale è cos`tuita da:

- Un insieme di nodi (neuroni), o unità connesse da collegamen`.
- Un insieme di pesi associa` ai collegamen`.
- Un insieme di soglie o livelli di agvazione.



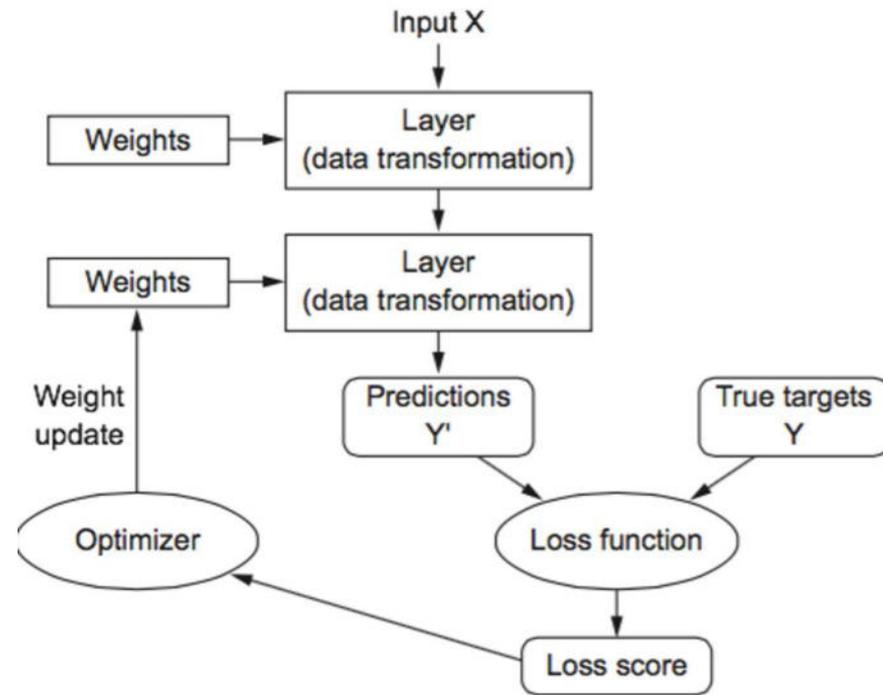
- Layers che sono combinati in una rete (o modello)
- I dati di input e gli obiettivi corrispondenti (*targets*)
- La funzione di costo (*loss function*), che definisce il segnale di retroazione utilizzato per l'apprendimento
- L'ottimizzatore (*optimizer*), che determina come procede l'apprendimento



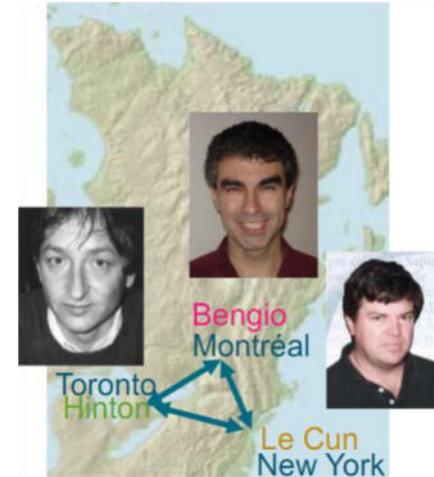
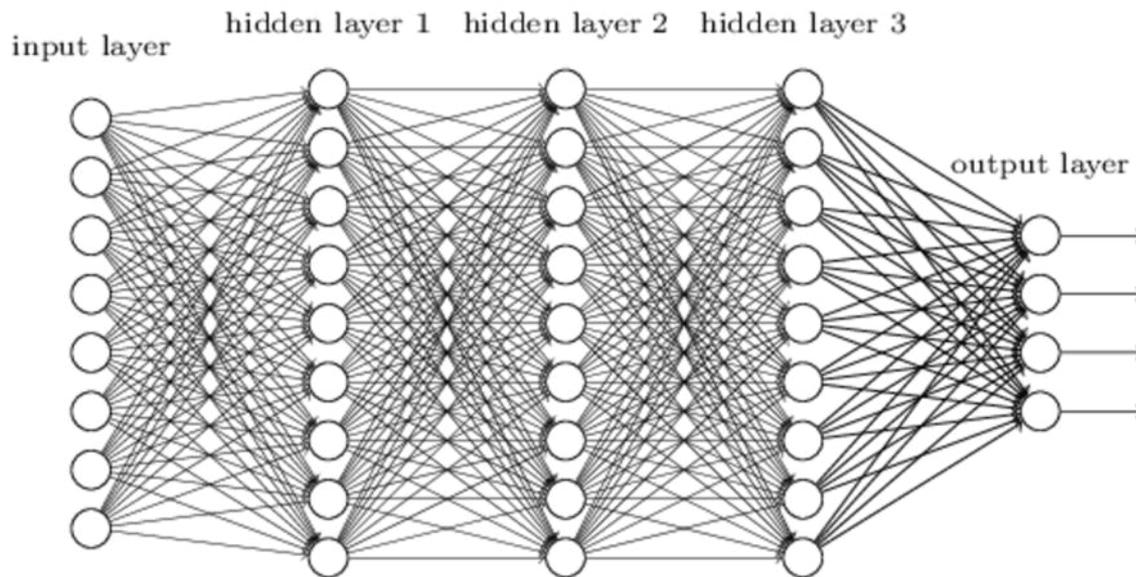
I pesi sono fondamentali perché determinano quanto ciascun input influenzi l'output di un neurone o di un'intera rete. Un peso maggiore indica che l'input corrispondente è considerato più importante, mentre un peso minore indica che l'input è meno rilevante.

I pesi consentono alla rete di apprendere e catturare le relazioni complesse tra i dati di input e gli output desiderati.

Durante la fase di addestramento, i pesi vengono regolarmente aggiornati in base all'errore tra gli output generati dalla rete e gli output desiderati. Questo processo è noto come algoritmo di apprendimento e mira a minimizzare l'errore della rete. I pesi vengono aggiornati in modo tale da ridurre l'errore e migliorare la capacità della rete di generalizzare e fare previsioni accurate su nuovi dati.



Il Deep learning è una tipologia di rete neurale profonda. La definizione di deep è associata al numero di strati nascosti presenti che contengono molti nodi con una moltitudine di connessioni.



Ricercatori premia` con il premio **ACM Turing Award 2019** per agvità di ricerca su back propaga`on, CCN, GAN, high-dimensional word embedding and aken`on, etc..

Intuizione di una rete neurali convoluzionale

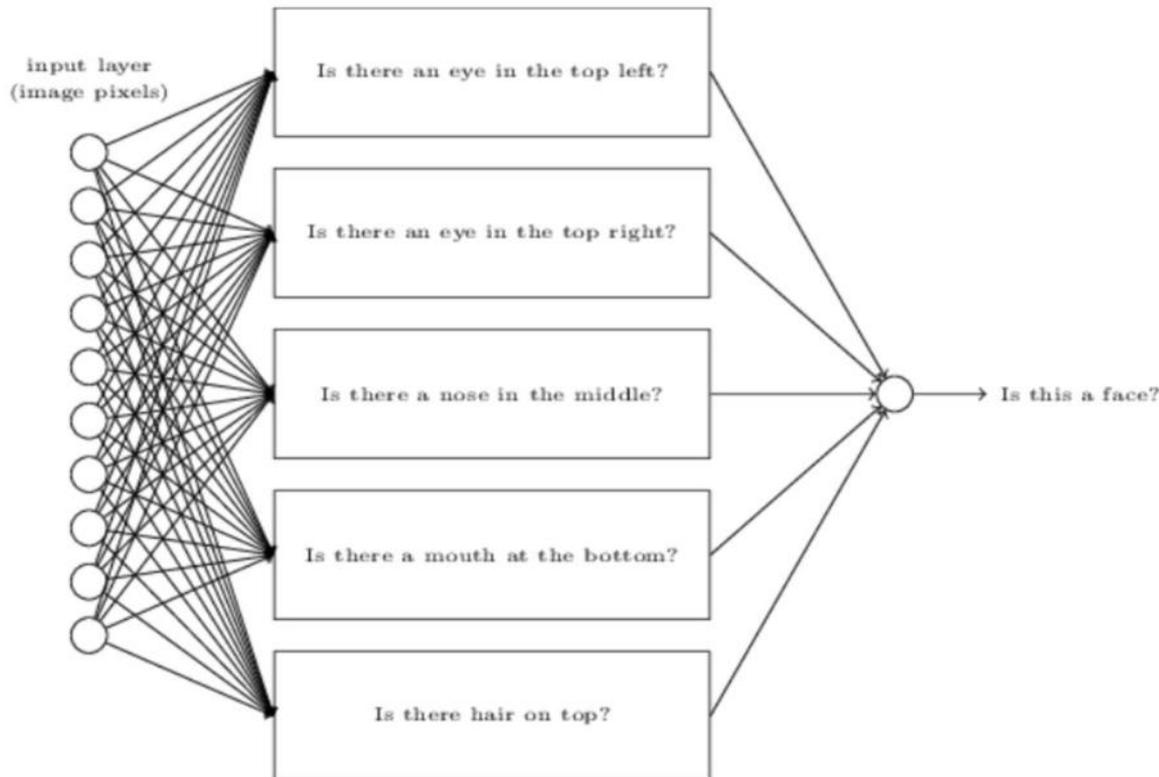
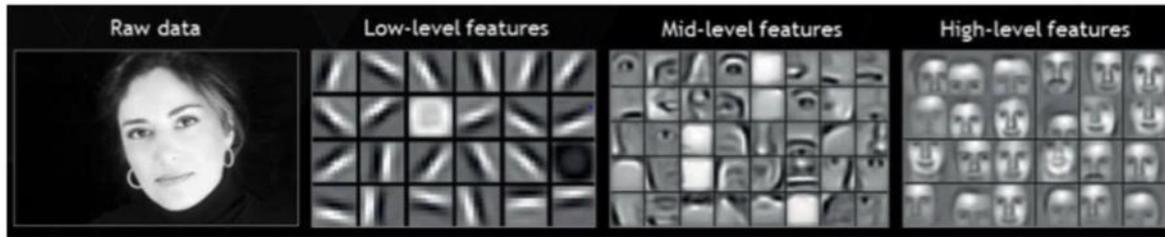
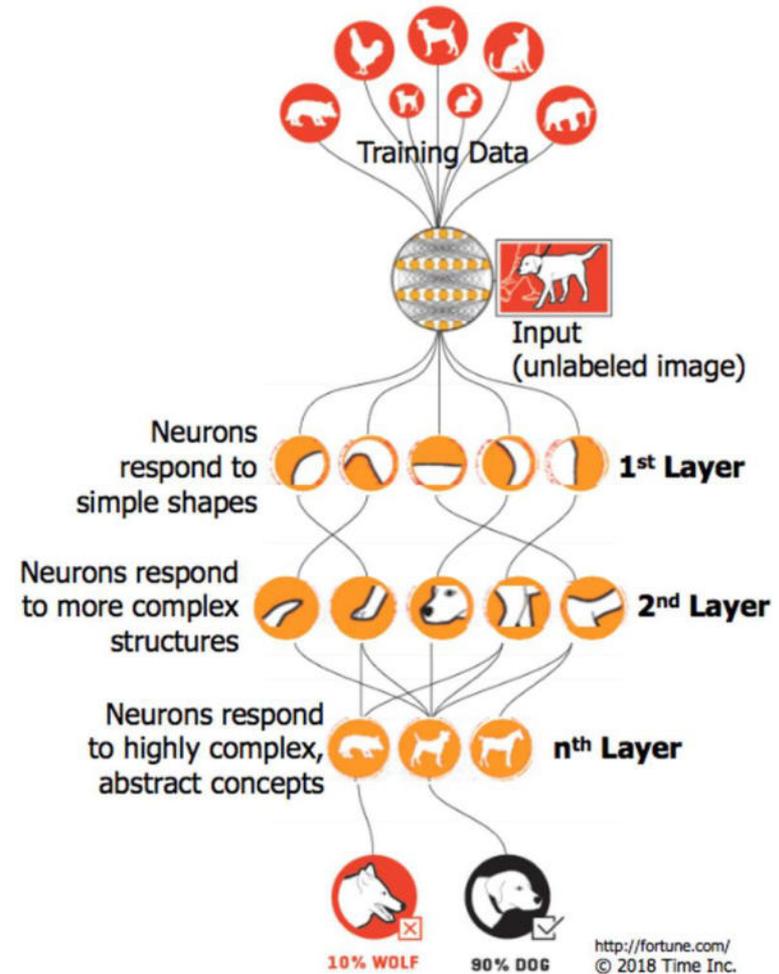
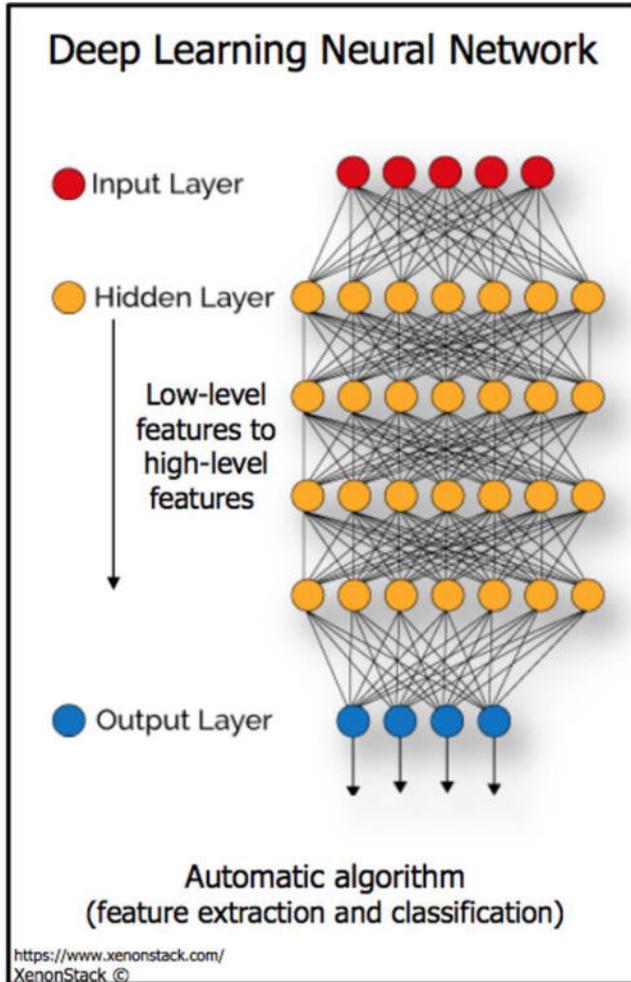


Illustration from Neural Networks and Deep Learning (Nielsen, 2017)



Deep learning - neural style

Estrarre delle caratteristiche e dei pattern da un'immagine e trasferirle ad un'altra



Notte stellata di Vincent Van Gogh + foto del campus di Stanford

Fonte: A Neural Algorithm of Artistic Style by Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge, 2015. <https://github.com/jcjohnson/neural-style>

Gen Studio è una collaborazione del Met, Microso{ e MIT per esplorare visualmente il tema della generazione di immagini di opere d'arte.

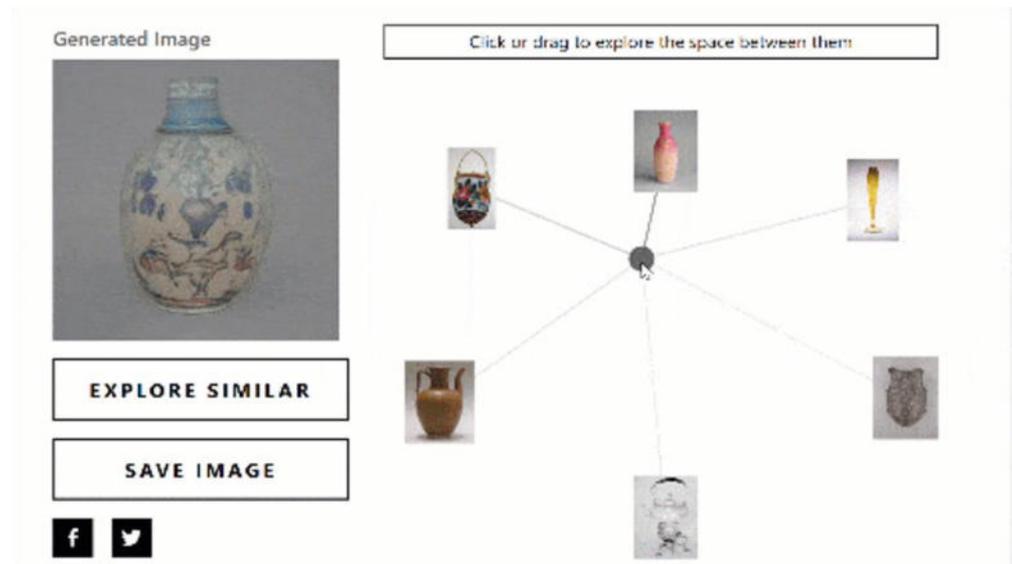
In par`colare Gen Studio permeke di:

- Estrarre la strukura delle opere d'arte presen` al MET;
- Interpolare tra le carakeris`che delle opere per generare nuovi oggeg.

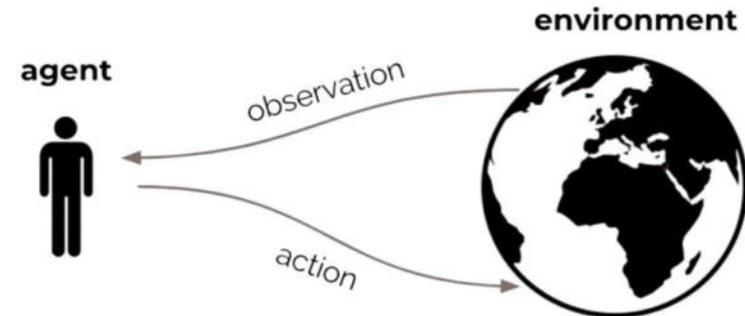
Fonte:

<https://github.com/Microsoft/GenStudio>

<https://www.metmuseum.org/blogs/now-at-the-met/2019/wikipedia-art-and-ai>



- Noi, e gli altri essere viventi, impariamo dall'interazione con il nostro ambiente
- Questa modalità è diversa rispetto ad altre tipologie di learning
 - Attiva e non passiva
 - Interazione è sequenziale, le future interazioni possono dipendere da quelle precedenti
- Nella nostra interazione con l'ambiente siamo focalizzati sugli obiettivi (goal directed)
- Possiamo imparare sbagliando
(Es: imparare ad andare in bicicletta)



Spunto: Deep mind, <https://www.youtube.com/channel/UCP7jMXY2xbc3KCAE0MHQ-A>

Reinforcement learning

Che cos'è



Quando usarlo

Una classe di algoritmi che imparano a svolgere un compito semplicemente cercando di massimizzare i premi che riceve per le sue azioni (Es: massimizzare i premi ricevuti per i rendimenti crescenti di un portafoglio di investimenti).

Non sono disponibili molti dati di training e non è possibile definire chiaramente quale è l'output finale ideale. L'unico modo per conoscere l'ambiente è attraverso l'interazione con esso.

Esempio

L'obiettivo è chiaramente vincere la partita, mentre le "situazioni" sono le varie configurazioni (ad esempio della scacchiera) e le "azioni" sono le mosse valide. Queste tecniche, fondamentalmente andando per tentativi, imparano un comportamento che lo porta con successo all'obiettivo.

Caratteristiche

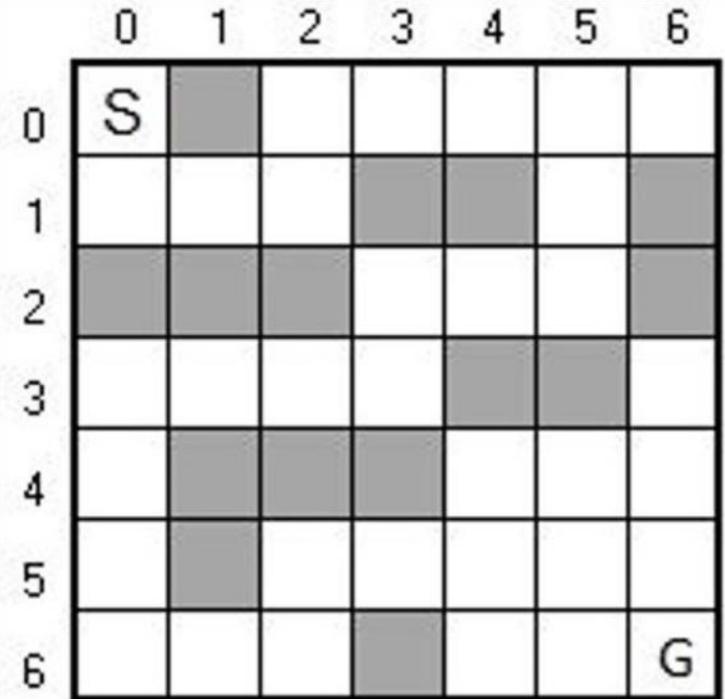
- **Un ambiente inizialmente sconosciuto:** è necessario esplorare l'ambiente per ottenere informazioni su di esso;
- **Feedback ritardato:** dopo aver intrapreso un'azione, potrebbe essere necessario del tempo per realizzare pienamente le conseguenze a lungo termine di tale azione;
- **Processo decisionale sequenziale:** la ricompensa complessiva ricevuta è in genere il risultato di una sequenza di più azioni piuttosto che una singola azione indipendente.



Esempio: un robot deve uscire da un labirinto

Per incoraggiare il robot a trovare il percorso più breve, viene applicata una piccola penalità di 0,04 unità ogni volta che il robot si sposta in una cella vuota (bianca) e gli ostacoli sono posti intorno al labirinto (contrassegnati in grigio) che provocano una penalità maggiore di 0,75 unità se il robot entra in una cella contenente una di esse.

Il robot può solo muoversi verso l'alto, il basso, a sinistra o a destra (ovvero, le mosse diagonali non sono consentite). Tuttavia, un livello di incertezza è associato a ciascun movimento, in modo tale che vi sia solo un 80% di probabilità che il robot si muova nella direzione desiderata e un 20% di probabilità che il robot si sposti ad angolo retto rispetto alla direzione desiderata.



Reinforcement learning - esempio

Il diagramma mostra la direzione ottimale per il robot da prendere in ogni quadrato della griglia.

	0	1	2	3	4	5	6
0	↓	→	→	→	→	↓	←
1	→	→	↑	↓	↓	↓	←
2	↓	↓	↓	↓	←	←	↓
3	↓	←	←	←	↓	↓	↓
4	↓	↑	↓	↓	→	↓	↓
5	↓	→	→	→	→	↓	↓
6	→	→	↑	↑	→	→	↓

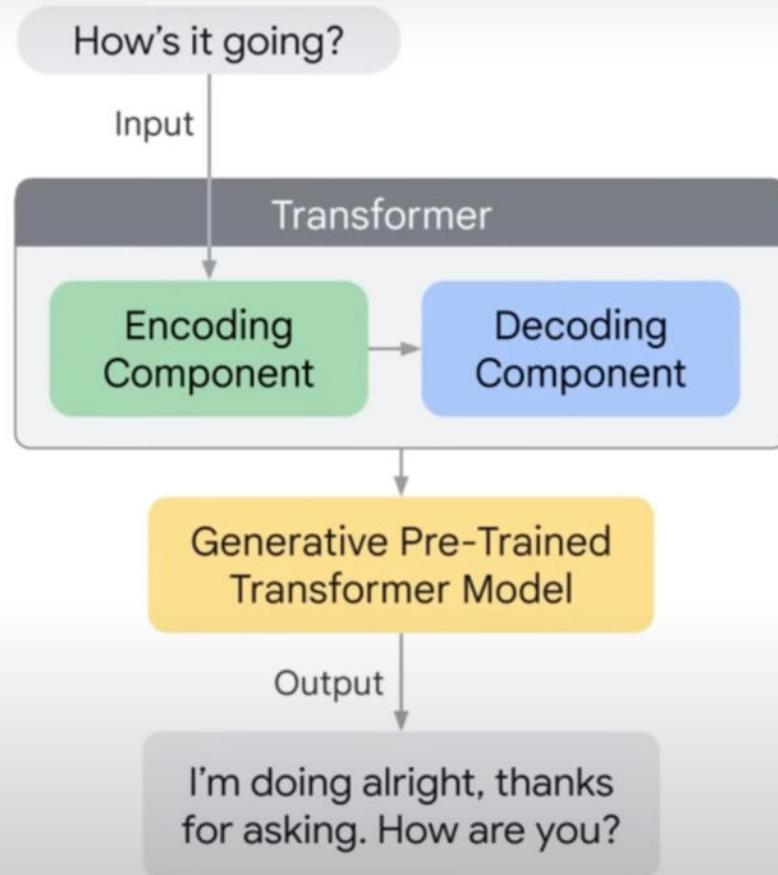
Ritorniamo a ChatGPT e ai Large Language Model

ChatGPT è un modello di linguaggio sviluppato da OpenAI messo a punto con **tecniche di machine learning non supervisionato e ottimizzato con tecniche di apprendimento supervisionato e per rinforzo.**

How it Works

Pre-Training:

- Large amount of Data
- Billions of parameters
- Unsupervised learning



Fonte: Google Cloud Tech

- Come esseri umani, percepiamo il testo come una raccolta/sequenza di parole. Le frasi sono sequenze di parole. I documenti sono sequenze di capitoli, sezioni e paragrafi. Tuttavia, per i computer, il testo è semplicemente una sequenza di caratteri.

```
Example of input text
As she said this, she looked down at her hands, and was surprised to find that she had put on one of the rabbit's little gloves while she was talking. "How can I have done that?" thought she, "I must be growing small again." She got up and went to the table to measure herself by it, and found that, as nearly as she could guess, she was now about two feet high, and was going on shrinking rapidly: soon she found out that the reason of it was the nosegay she held in her hand: she dropped it hastily, just in time to save herself from shrinking away altogether, and found that she was now only three inches high.
```

```
Tokenized Text
['As', 'she', 'said', 'this', ',', 'she', 'looked', 'down', 'at', 'her', 'hands', ',', 'and', 'was', 'surprised', 'to', 'find', 'that', 'she', 'had', 'put', 'on', 'one', 'of', 'the', 'rabbit', "'s", 'little', 'gloves', 'while', 'she', 'was', 'talking', ',', 'I', 'must', 'be', 'growing', 'small', 'again', '."', 'She', 'got', 'up', 'and', 'went', 'to', 'the', 'table', 'to', 'measure', 'herself', 'by', 'it', 'and', 'found', 'that', 'as', 'nearly', 'as', 'she', 'could', 'guess', 'she', 'was', 'now', 'about', 'two', 'feet', 'high', 'and', 'was', 'going', 'on', 'shrinking', 'rapidly', ':', 'soon', 'she', 'found', 'out', 'that', 'the', 'reason', 'of', 'it', 'was', 'the', 'nose', 'gay', 'she', 'held', 'in', 'her', 'hand', ':', 'she', 'dropped', 'it', 'hastily', ',', 'just', 'in', 'time', 'to', 'save', 'herself', 'from', 'shrinking', 'away', 'altogether', 'and', 'found', 'that', 'she', 'was', 'now', 'only', 'three', 'inches', 'high', '.']
```

```
Embedding of the Above Text
[ 2.49 0.22 -0.36 -1.55 0.22 -2.45 2.65 -1.6 -0.14 2.26
-1.26 -0.61 -0.61 -1.89 -1.87 -0.16 3.34 -2.67 0.42 -1.71
...
2.91 -0.77 0.13 -0.24 0.63 -0.26 2.47 -1.22 -1.67 1.63
1.13 0.03 -0.68 0.8 1.88 3.05 -0.82 0.09 0.48 0.33]
```

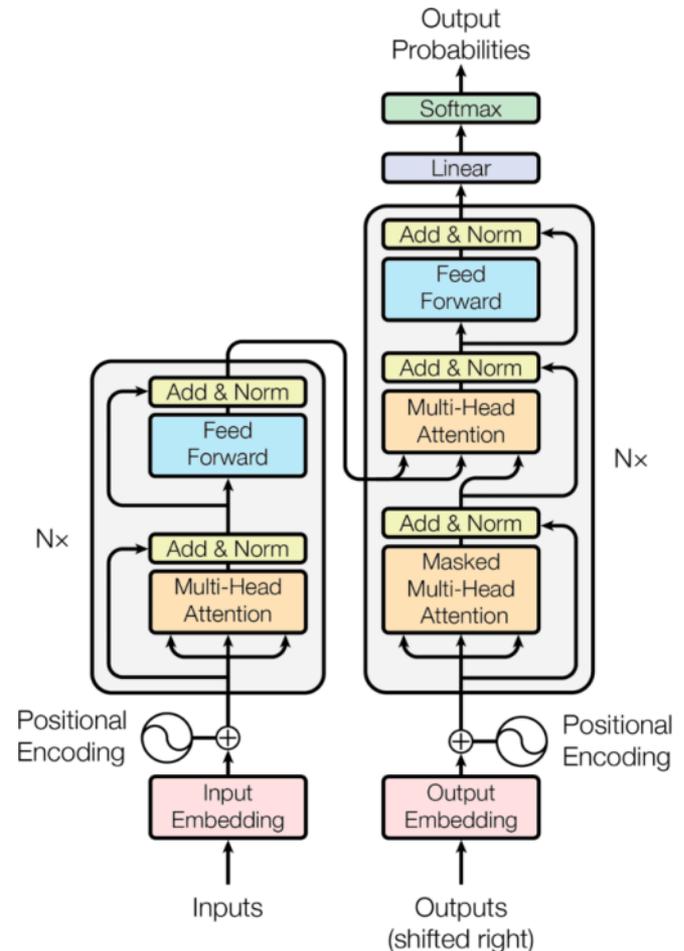
Nel 2017 Vaswani et al. ha pubblicato un documento, "Attention is all you need", per introdurre un modello di **Transformer**. Si basa sul meccanismo dell'attenzione. Contrariamente alle reti neurali ricorrenti, il meccanismo dell'attenzione consente di vedere l'intera frase (o anche il paragrafo) in una volta anziché una parola alla volta, **per dare ancora più senso alle lunghe sequenze linguistiche, creando connessioni tra diverse parti della sequenza.** Ciò consente al modello del trasformatore di comprendere meglio il contesto di una parola.

Attention is all you need

[A Vaswani, N Shazeer, N Parmar...](#) - Advances in neural ..., 2017 - proceedings.neurips.cc

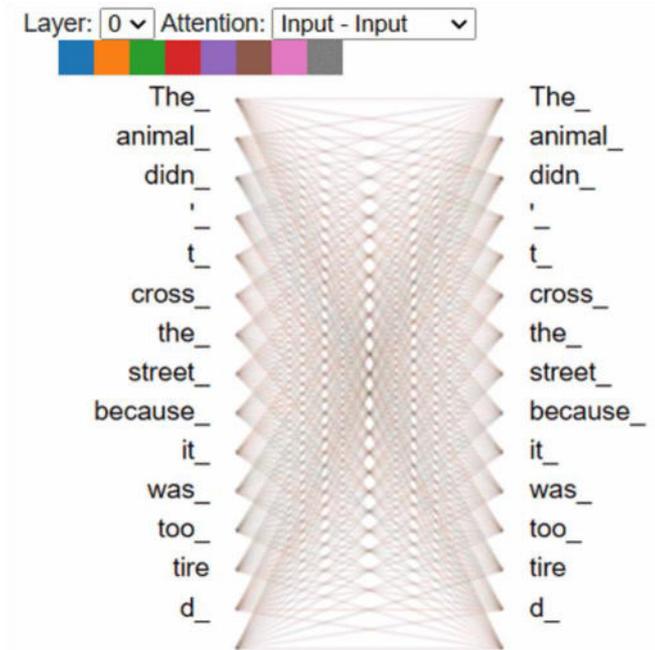
... to attend to **all** positions in the decoder up to and including that position. **We need** to prevent ... **We** implement this inside of scaled dot-product **attention** by masking out (setting to $-\infty$) ...

☆ Salva 📄 Cita Citato da 76521 Articoli correlati Tutte e 46 le versioni 🔗



Transformers (3)

- **i transformers intervengono per catturare le relazioni tra i token e l'ordine in cui compaiono nella frase.** I transformers utilizzano il meccanismo di attenzione per dare più peso a determinati token in base al contesto e alle relazioni con gli altri token.
- I pesi nella rete neurale rappresentano i parametri che determinano l'importanza relativa dei segnali di input. Durante l'addestramento, i pesi vengono aggiornati iterativamente in base all'errore tra l'output generato dal modello e l'output desiderato, al fine di migliorare le prestazioni della rete.
- **GPT - Generative Pre-trained Transformer**
- **BERT - Bidirectional Encoder Representations from Transformers**



Fonte:

https://colab.research.google.com/github/tensorflow/tensor2tensor/blob/master/tensor2tensor/notebooks/hello_t2t.ipynb#scrollTo=OJKU36QAfqOC

Procedimento

- Addestramento non supervisionato su una dataset di testo di grandi dimensioni
- Addestramento supervisionato (fine tuning) su un dataset specializzato con da ` di input e output per un task specifico

Modelli pre-trained

L'addestramento di un modello richiede risorse importan` a livello di dataset e potenza computazionale. Si possono anche usare modelli pre addestra` (Huggingfaces, Bert di Google, I modelli GPT-3 di OpenAI)

Apprendimento per rinforzo dal feedback umano (RLHF)

L'apprendimento per rinforzo ordinario, in cui gli agenti apprendono dalle proprie azioni sulla base di una "funzione di ricompensa", è difficile da applicare alle attività di elaborazione del linguaggio naturale perché le ricompense spesso non sono facili da definire o misurare.

Come si crea / codifica una loss function per definire **concetti** come: **divertente, etico o sicuro** ?

Inoltre, i modelli di linguaggio di grandi dimensioni possono generare **output che sono falsi, tossici o semplicemente non utili per l'utente**. In altre parole, questi i modelli non sono allineati con i loro utenti.

Open AI nel 2017 ha sviluppato un algoritmo di apprendimento per rinforzo dal feedback umano (RLHF) è una tecnica che addestra un "**modello di ricompensa**" **direttamente dal feedback umano e utilizza il modello come funzione di ricompensa per ottimizzare la politica di un agente utilizzando l'apprendimento per rinforzo (RL)** attraverso un algoritmo di ottimizzazione come Proximal Policy Optimization. Il feedback umano viene raccolto chiedendo agli umani di classificare le istanze del comportamento dell'agente.

Fonte:

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., ... & Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35, 27730-27744.

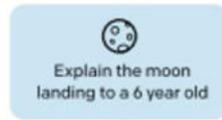
Schulman, John; Wolski, Filip; Dhariwal, Prafulla; Radford, Alec; Klimov, Oleg (2017). "Proximal Policy Optimization Algorithms".

Supervised fine tuning - Reinforcement learning from human feedback

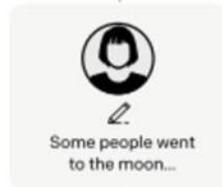
Step 1

Collect demonstration data, and train a supervised policy.

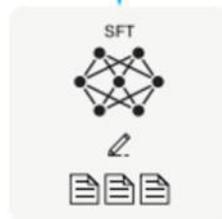
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



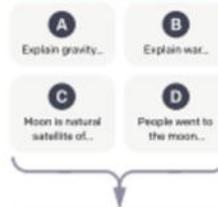
This data is used to fine-tune GPT-3 with supervised learning.



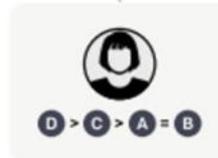
Step 2

Collect comparison data, and train a reward model.

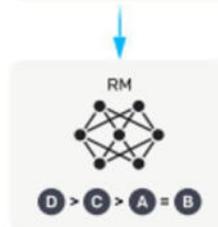
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



Step 3

Optimize a policy against the reward model using reinforcement learning.

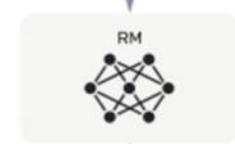
A new prompt is sampled from the dataset.



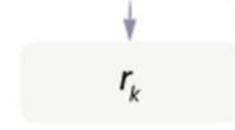
The policy generates an output.



The reward model calculates a reward for the output.



The reward is used to update the policy using PPO.



Fonte: Google Cloud Tech

ULTERIORI SPUNTI

Autonomous AI

Autonomous agents

Come modello di linguaggio AI, ChatGPT può essere considerato un agente intelligente.

Un agente intelligente è un sistema in grado di percepire il suo ambiente, ragionarci sopra e intraprendere azioni basate sulla sua comprensione per raggiungere scopi o obiettivi specifici.

ChatGPT può percepire ed elaborare input in linguaggio naturale, ragionare su di essi in base ai suoi numerosi dati di addestramento e generare risposte appropriate o eseguire attività specifiche.

Tuttavia, **l'autonomia di ChatGPT è limitata dalla sua stretta attenzione all'elaborazione e alla comprensione del linguaggio.** Può solo "prendere decisioni" nell'ambito della generazione di testo pertinente in base all'input e **non possiede la capacità di adattarsi o apprendere in modo indipendente oltre la sua formazione.**

Gli agenti autonomi non sono semplicemente più intelligenti dei modelli di base su cui si basano, ma aprono una dimensione completamente nuova: sono capaci di "pensare lentamente". Risolvono domande complicate, in cui si giunge all'obiettivo poco a poco tramite risultati intermedi.

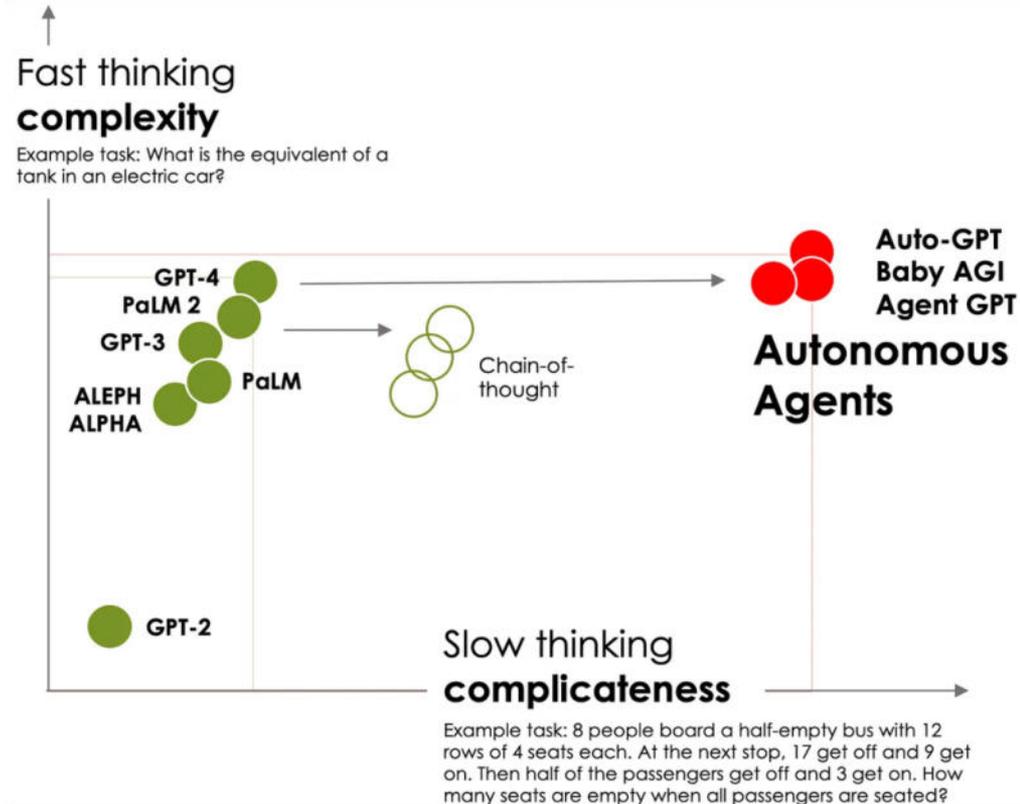
Tra le funzionalità aggiuntive:

- navigare su internet
- operazione di lettura/scrittura
- interagire con altri agenti GPT
- eseguire codice
- ..

Fonte:

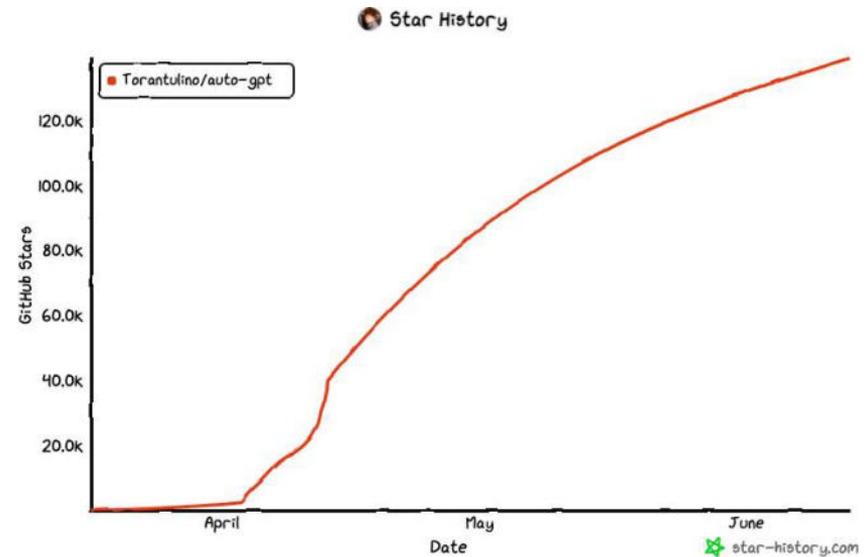
Shen, Y., Song, K., Tan, X., Li, D., Lu, W., & Zhuang, Y. (2023). Hugginggpt: Solving ai tasks with chatgpt and its friends in huggingface. *arXiv preprint arXiv:2303.17580*.

Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. *arXiv preprint arXiv:2304.03442*



Auto-GPT è un agente intelligente capace di scomporre un obiettivo assegnatogli in linguaggio naturale e di dedurre delle sotto-attività più semplici che esegue mediante Internet e altri strumenti all'interno di un'iterazione automatica.

A differenza dei sistemi interattivi come ChatGPT, che necessitano dell'inserimento di comandi manuali per ogni attività, Auto-GPT si assegna nuovi sotto-obiettivi funzionali al conseguimento di un obiettivo maggiore, senza la necessità di un costante input da parte umana e così facendo crea e modifica i propri prompt con istanze ricorsive e in risposta alle nuove informazioni acquisite.



Fonte: <https://github.com/Significant-Gravitas/Auto-GPT>

Features:

- Accesses the internet for queries and gathering information
- Long and short-term memory management
- GPT-4 instances for text generation
- Accesses popular websites and platforms
- File storage and summarization with GPT-3.5

★ **Embarking on a new goal:** Create a timeline and checklist for organizing a corporate conference for 200 attendees in Rome at the end of June. Please provide a list of 3 available hotels for hosting the conference with the price of rooms and catering

🟡 **Thinking...** (Restart if this takes more than 30 seconds)

★ **Added task:** Determine the conference dates and location, ensuring availability and accessibility for attendees.

★ **Added task:** Choose 3 potential hotels for hosting the conference, considering factors such as price, amenities, and catering options.

★ **Added task:** Contact the hotels to obtain room and catering pricing for the estimated number of attendees.

★ **Added task:** Create a timeline and checklist for organizing the conference, including tasks such as booking hotel rooms, arranging travel, and planning event activities.

🟡 **Thinking...** (Restart if this takes more than 30 seconds)

Fonte: <https://agentgpt.reworkd.ai/>

AgentGPT (2)

First, we need to identify available hotels for hosting the conference. After researching and considering factors such as location, amenities, and pricing, I recommend the following three options:

1. Grand Hotel Palace Rome – This hotel is located close to major attractions in Rome and offers a variety of event spaces suitable for hosting a conference. The price for the conference room is €3,000 per day, and catering costs €50 per person.
2. Hotel Bernini Bristol – This hotel is situated in the center of Rome and also offers several event spaces for hosting conferences. The price for renting the conference room is €2,500 per day, and catering costs €45 per person.
3. Parco dei Principi Grand Hotel & SPA – This hotel is 20 minutes away from the city center, but offers a spacious conference room that can accommodate up to 300 attendees. The price for the conference room is €2,800 per day, and catering costs €55 per person.

Next, we need to create a timeline and checklist for organizing the conference. This timeline should be based on the specific details of the event, and should include tasks such as:

- Booking hotel rooms for attendees and speakers
- Arranging travel and transportation for attendees
- Planning the conference agenda and selecting speakers
- Coordinating catering and other event logistics
- Promoting the conference and managing registrations

To ensure that these tasks are completed efficiently and effectively, it may be helpful to assign specific responsibilities to different members of the team or to work with an event planner. By carefully considering all of the necessary steps and making strategic decisions along the way, we can create a successful and memorable conference in Rome for all attendees.

Fonte: <https://agentgpt.reworkd.ai/>

25 Maggio
2023

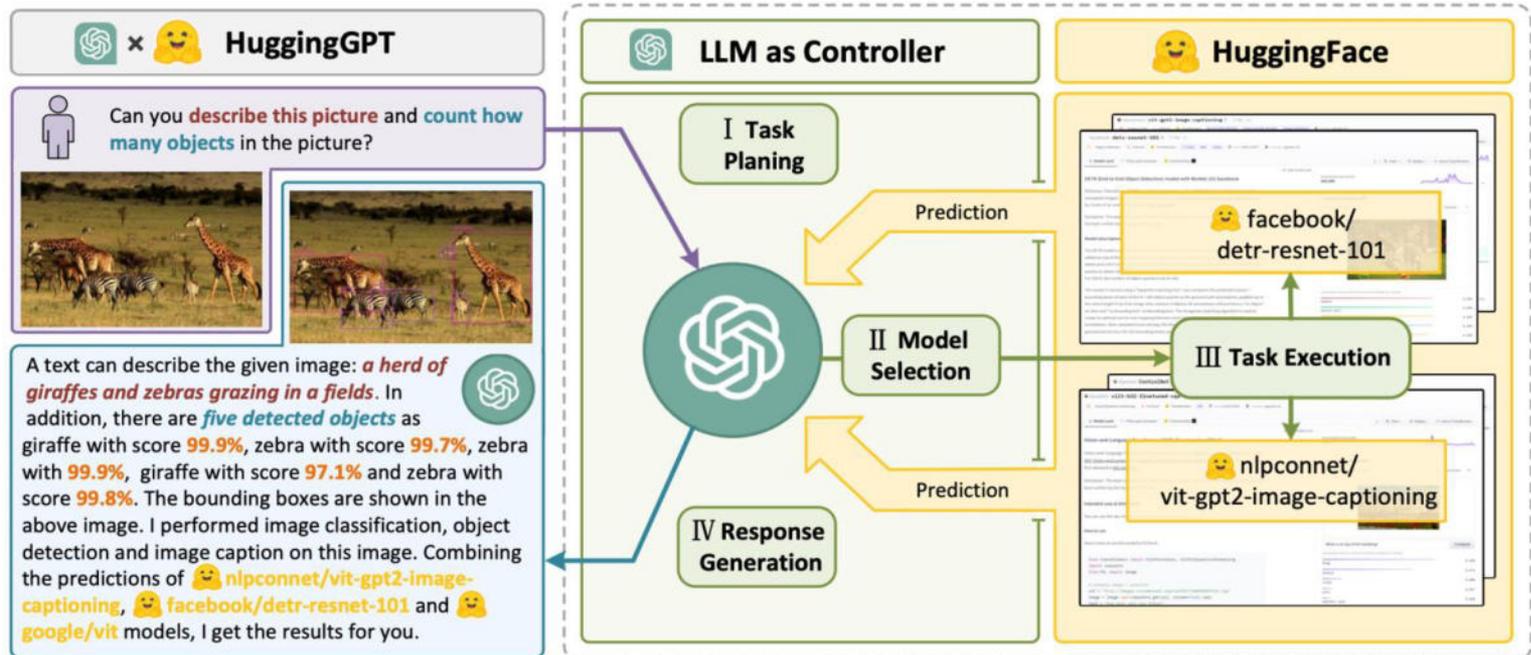
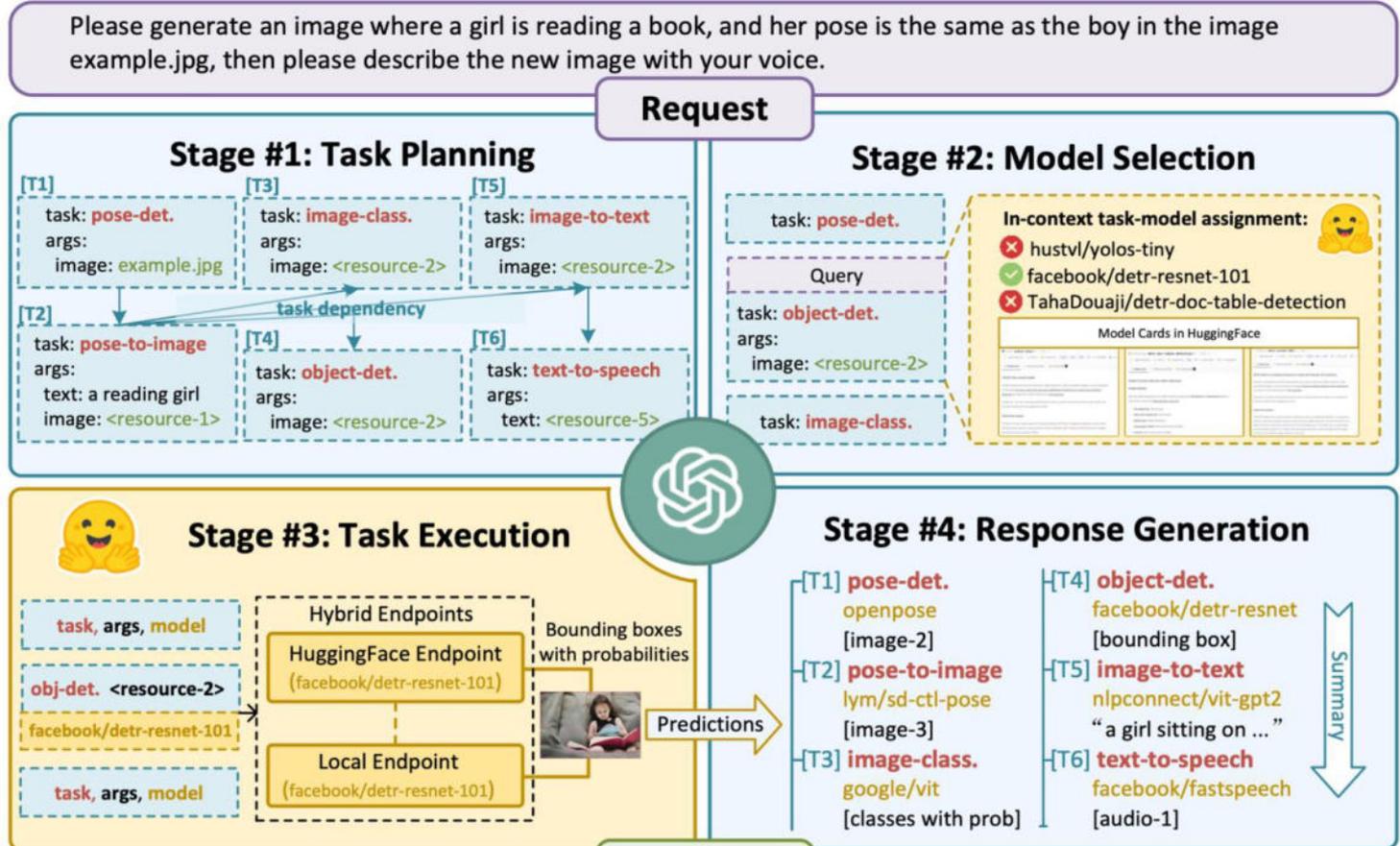


Figure 1: Language serves as an interface for LLMs (e.g., ChatGPT) to connect numerous AI models (e.g., those in Hugging Face) for solving complicated AI tasks. In this concept, an LLM acts as a controller, managing and organizing the cooperation of expert models. The LLM first plans a list of tasks based on the user request and then assigns expert models to each task. After the experts execute the tasks, the LLM collects the results and responds to the user.

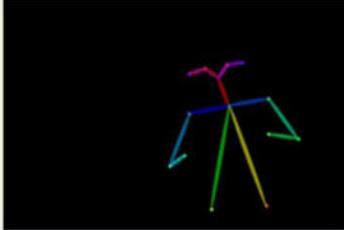
Fonte: Shen, Y., Song, K., Tan, X., Li, D., Lu, W., & Zhuang, Y. (2023). Hugginggpt: Solving ai tasks with chatgpt and its friends in huggingface. *arXiv preprint arXiv:2303.17580*

25 Maggio
2023



Fonte: Shen, Y., Song, K., Tan, X., Li, D., Lu, W., & Zhuang, Y. (2023). Hugginggpt: Solving ai tasks with chatgpt and its friends in huggingface. *arXiv preprint arXiv:2303.17580*

Response



a girl sitting on a bed reading a book

[Text-1] = <resource-5>



[Audio-1]

[Image-1] = example.jpg [Image-2] = <resource-1> [Image-3] = <resource-2>

[Image-4]

The image you gave me is of "boy". The first thing I did was use 🤖 [OpenCV's openpose model](#) to analyze the pose of the boy in the image. I fed that image, specifically located at [Image-1], into the model to get an output image of the pose, located at [Image-2]. Then I take that pose and generate a new image, located at [Image-3], using 🤖 [llyasviel/sd-controlnet-openpose](#) and the instruction "a girl is reading". Next, I ran object detection on the image generated in the previous step, to get predictions of the objects present in the image, including the box location. These predictions were run through the 🤖 [facebook/detr-resnet-101](#) model. The box was drawn on the image, which is located at [Image-4]. Then, I used the predictions of a image classification model 🤖 [google/vit-base-patch16-224](#) and a image caption model a 🤖 [nlpconnect/vit-gpt2-image-captionin](#) model to generate the caption for newly generated image. It generated text: "a girl sitting on a bed reading a book". Lastly, I ran the 🤖 [facebook/fastspeech2-en-ljspeech](#) model to convert the generated text into audio, located at [Audio-1].

Fonte: Shen, Y., Song, K., Tan, X., Li, D., Lu, W., & Zhuang, Y. (2023). Hugginggpt: Solving ai tasks with chatgpt and its friends in huggingface. *arXiv preprint arXiv:2303.17580*

Normativa e attività della UE

L'UE ha supportato l'intelligenza artificiale per molti anni, in gran parte attraverso i suoi programmi di ricerca e sviluppo, 7° PQ e Horizon 2020.

42 nuovi progetti su AI e la robotica in questi campi:

INNOVATION IN AI, DATA AND ROBOTICS

AI, Data and Robotics for the Green Deal

INNOVATION IN AI, DATA AND ROBOTICS

AI, Data and Robotics at work

LEADERSHIP IN AI BASED ON TRUST

Tackling gender, race and other biases in AI

LEADERSHIP IN AI BASED ON TRUST

Verifiable robustness, energy efficiency and transparency for Trustworthy AI:
Scientific excellence boosting industrial competitiveness

TOMORROW'S DEPLOYABLE ROBOTS: EFFICIENT, ROBUST, SAFE, ADAPTIVE AND TRUSTED



Fonte: <https://digital-strategy.ec.europa.eu/en/library/horizon-europe-new-projects-robo4cs-and-ai-june-november-2022>

“L'Intelligenza Artificiale, oggi, può guidare al posto nostro, prendersi cura delle persone anziane o malate, svolgere lavori pericolosi o usuranti, aiutarci a prendere decisioni ponderate, basate sulla gestione razionale di grandi moli di dati. Ci può permettere di comunicare in lingue che non conosciamo, può seguirci nello studio e aumentare le esperienze culturali o di intrattenimento a nostra disposizione.

Nella PA può essere utilizzata con profitto nel sistema sanitario, scolastico, giudiziario, nel pubblico impiego, nella sicurezza e, più in generale, nella gestione delle relazioni coi cittadini, che possono venire semplificate e rese allo stesso tempo più efficaci, veloci ed efficienti. “

Fonte: Libro Bianco sull'Intelligenza Artificiale al servizio del cittadino, 2018, Task force sull'Intelligenza Artificiale dell'Agenzia per l'Italia Digitale - <https://ia.italia.it/assets/librobianco.pdf>

ENTRO IL
2025
L'INTELLIGENZA ARTIFICIALE
POTREBBE

MOLTO PROBABILMENTE

Incidere una canzone che arrivi nella Top 100 di Billboard
Creare un'opera d'arte valutata 100.000\$
Scrivere una serie TV di successo

POCO PROBABILMENTE

Firmare un best seller per il New York Times
Scrivere un film da Oscar
Vincere il premio Pulitzer per il giornalismo

Fonte: AI takes center stage, Bot.Me: A revolutionary partnership, PwC 2017

QUALI POSSONO ESSERE I POTENZIALI DANNI ASSOCIATI A SISTEMI IA

DIGITALI	FISICI	POLITICI
<ul style="list-style-type: none"> ✓ Phishing automatizzato, creazione di email false, siti web e link per sottrarre informazioni. 	<ul style="list-style-type: none"> ✓ Terrorismo automatizzato che utilizza droni o veicoli autonomi come armi. 	<ul style="list-style-type: none"> ✓ Propaganda, attraverso immagini e video falsi facilmente generabili.
<ul style="list-style-type: none"> ✓ Hacking massivo, attraverso la scoperta automatizzata delle vulnerabilità dei sistemi. 	<ul style="list-style-type: none"> ✓ Sciame di Robot che cercano di raggiungere lo stesso obiettivo. 	<ul style="list-style-type: none"> ✓ Rimozione automatica del dissenso, grazie alla possibilità di trovare, analizzare e rimuovere automaticamente testi ed immagini.
<ul style="list-style-type: none"> ✓ Ingannare sistemi IA sfruttando a proprio vantaggio i difetti nell'interpretazione della realtà da parte dell'IA. 	<ul style="list-style-type: none"> ✓ Attacchi remoti, resi possibili dal fatto che i robot autonomi possono essere controllati da qualunque distanza. 	<ul style="list-style-type: none"> ✓ Persuasione personalizzata, con l'utilizzo di informazioni pubbliche per influenzare l'opinione di qualcuno.

Fonte: AI experts list the real dangers of artificial intelligence, Dave Gershgorn Feb. 2018

Fonte: Libro Bianco sull'Intelligenza Artificiale al servizio del cittadino, 2018, Task force sull'Intelligenza Artificiale dell'Agenzia per l'Italia Digitale - [hGps://ia.italia.it/assets/librobianco.pdf](https://ia.italia.it/assets/librobianco.pdf)

- **Etica** – responsabilità, qualità e neutralità dei dati
- **Tecnologica** - creazione di algoritmi sempre più efficaci
- **Competenze** – skills necessari a sviluppare soluzioni di IA
- **Dati** – qualità ed organizzazione
- **Contesto legale** – garantire trasparenza
- **Bilanciamento interessi collettività e individui** – trasparenza atti amministrativi, protezione dati personali e accountability
- **Accompagnare la trasformazione del paese** – azioni necessarie per favorire utilizzo IA nel rapporto tra Stato, cittadini e imprese
- **Prevenire le diseguaglianze** – accessibilità per tutti e rimozione del bias
- **Misurazione dell'impatto** – metriche di customer satisfaction / ottimizzazione dei processi
- **L'essere umano** – strumenti per facilitare la comprensione ed il coinvolgimento delle persone

Fonte: Libro Bianco sull'Intelligenza Artificiale al servizio del cittadino, 2018, Task force sull'Intelligenza Artificiale dell'Agenzia per l'Italia Digitale - <https://ia.italia.it/assets/librobianco.pdf>

Il GDPR afferma: ***"L'interessato ha il diritto di non essere soggetto a una decisione basata esclusivamente sull'elaborazione automatica."***

La chiave qui è la parola "esclusivamente", il che significa che se un'organizzazione vuole utilizzare un processo decisionale automatico (per l'elaborazione delle richieste di prestito o degli adeguamenti salariali dei dipendenti), un umano deve esaminare e confermare le conclusioni raggiunte da un algoritmo IA prima di eseguirle.

Una persona può controllare la decisione automatica di una IA solo se è in grado di rintracciare i passaggi che l'hanno originata.

Strategico / finanziario:

Le soluzioni basate su IA sono affidabili per essere inserite nei prodotti e nei modelli di business ? Possiamo delegare decisioni ad un sistema di intelligenza artificiale?

Reputazionale:

Quale è l'impatto nel caso in cui una soluzione di IA sfugga al controllo ?

Legale / normativo:

La nostra soluzione di IA è conforme alla legge e possiamo dimostrare come si sono originate le sue decisioni e le sue azioni ?

Etico:

L'IA rispetta degli standard di tipo etico ?

Implicazioni sociali:

Che impatto può avere una soluzione di IA sulla popolazione e sui posti di lavoro ?

La legge sui servizi digitali (DSA) e la legge sui mercati digitali (DMA) mirano a creare uno spazio digitale più sicuro in cui siano protetti i diritti fondamentali degli utenti e a creare condizioni di parità per le imprese. Il DSA introduce misure di trasparenza per migliorare la trasparenza dei sistemi di raccomandazione e della pubblicità.

- Ai sensi dell'Art. 14(1), i fornitori di servizi intermediari devono includere nelle loro condizioni generali informazioni sulle misure e gli strumenti utilizzati per la moderazione dei contenuti, comprese le decisioni algoritmiche. Inoltre indica come i destinatari del loro servizio siano adeguatamente informati su come gli algoritmi influenzano e condizionano il modo in cui le informazioni vengono visualizzate.
- Ai sensi dell'Art. 27(1), Nei sistemi di raccomandazione i fornitori devono specificare nelle loro condizioni generali "i principali parametri utilizzati nei loro sistemi di raccomandazione, nonché eventuali opzioni per i destinatari del servizio per modificare o influenzare tali parametri principali."

Fonte: [hTps://digital-strategy.ec.europa.eu/it/policies/digital-services-act-package](https://digital-strategy.ec.europa.eu/it/policies/digital-services-act-package)



The European Centre for Algorithmic Transparency (ECAT) will contribute to a safer, more predictable and trusted online environment for people and business.

How algorithmic systems shape the visibility and promotion of content, and its societal and ethical impact, is an area of growing concern. Measures adopted under the Digital Services Act (DSA) call for algorithmic accountability and transparency audits.

I membri del Parlamento Europeo mirano a garantire che i sistemi di intelligenza artificiale siano supervisionati da persone, siano sicuri, trasparenti, rintracciabili, non discriminatori ed ecologicamente sostenibili.

Vogliono anche avere una definizione uniforme per l'AI progettata per essere neutrale rispetto alla tecnologia, in modo che possa applicarsi ai sistemi di AI di oggi e di domani.

- Una volta approvate, sarà la prima regolamentazione al mondo sull'Intelligenza Artificiale.
- Prevede divieti per sorveglianza biometrica, riconoscimento delle emozioni, sistemi di AI per la polizia predittiva.
- Prevede regolamenti per l'AI a scopo generale e modelli di base come GPT.
- Prevede il diritto per i cittadini di presentare reclami riguardo ai sistemi di AI.

Fonte: **Parlamento europeo 11-05-2023** <https://www.europarl.europa.eu/news/it/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence>

I membri del Parlamento Europeo hanno incluso obblighi per i fornitori di modelli che dovranno:

- **garantire una protezione solida dei diritti fondamentali, della salute e della sicurezza, dell'ambiente, della democrazia e dello stato di diritto.**
- Dovranno valutare e mitigare i rischi, rispettare i requisiti di progettazione, informazione e ambiente e registrarsi nel database dell'UE.
- **I modelli di base generativi, come ChatGPT, dovranno rispettare ulteriori requisiti di trasparenza, come dichiarare che il contenuto è stato generato da AI, progettare il modello in modo da impedirgli di generare contenuti illegali e pubblicare riepiloghi dei dati protetti da copyright utilizzati per l'addestramento.**

AI ACT - Sostenere l'innovazione e proteggere i diritti dei cittadini

Per promuovere l'innovazione nell'AI, i membri del Parlamento Europeo hanno aggiunto eccezioni a queste regole per le attività di ricerca e per i componenti di AI forniti con licenze open-source.

La nuova legge promuove le aree di sperimentazione regolamentate, o ambienti controllati, istituiti dalle autorità pubbliche per testare l'AI prima della sua implementazione.

I membri del Parlamento Europeo desiderano **rafforzare il diritto dei cittadini di presentare reclami riguardanti i sistemi di AI e ricevere spiegazioni sulle decisioni basate su sistemi di AI ad alto rischio che influiscono significativamente sui loro diritti.**

AI ACT - Divieto di usi invasivi e discriminatori

- Sistemi di identificazione biometrica in tempo reale in spazi pubblicamente accessibili;
- Sistemi di identificazione biometrica a distanza, ad eccezione delle forze dell'ordine per perseguire crimini gravi e solo previa autorizzazione giudiziaria;
- Sistemi di categorizzazione biometrica che utilizzano caratteristiche sensibili (ad esempio, genere, razza, etnia, cittadinanza, religione, orientamento politico);
- Sistemi di polizia predittiva (basati sul profilo, sulla posizione o sul comportamento criminale passato);
- Sistemi di riconoscimento delle emozioni in applicazioni di forze dell'ordine, gestione delle frontiere, ambiente di lavoro ed istituti educativi;
- Estrazione indiscriminata di dati biometrici da social media o video di CCTV per creare database di riconoscimento facciale (violazione dei diritti umani e del diritto alla privacy).

[HOME](#) / [TECNOLOGIA](#)

ChatGPT come Google Bard? Parla Sam Altman: "pronti a ritirare l'IA dall'Unione Europea"

Il CEO di OpenAI, Sam Altman, ha avvertito che l'azienda potrebbe ritirare i suoi servizi dall'UE, se le future regole sulle IA saranno troppo restrittive.

Accesso prioritario ai modelli IA

Il Primo Ministro del Regno Unito ha annunciato la creazione di una taskforce di esperti e un investimento di 100 milioni di sterline per la ricerca sull'intelligenza artificiale, in particolare sulla sicurezza della tecnologia.

Sunak ha dichiarato che **Google DeepMind, OpenAI e Anthropic forniranno accesso in anteprima o prioritario ai loro modelli IA per scopi di ricerca**. Ciò consentirà di valutare e comprendere meglio le opportunità e i rischi di questi sistemi. Non è chiaro però quali dati verranno condivisi con il governo britannico. Tuttavia il Regno Unito potrebbe anticipare di molti mesi l'entrata in vigore dell'AI Act (proposto dalla Commissione europea) che prevede obblighi di trasparenza per le aziende, tra cui l'indicazione dei dati coperti dal copyright usati per l'addestramento.

Fonte: <https://www.punto-informa4co.it/regno-unito-accedera-modelli-ia-openai/> 13 giugno

Strategia di ENISA per la AI

L'Agencia dell'Unione Europea per la Cybersecurity (ENISA) ha pubblicato il 7 giugno 2023 4 report sulle sfide cyber riguardanti l'intelligenza artificiale:

- Multilayer Framework for Good Cybersecurity Practices for AI
- Cybersecurity and privacy in AI - Forecasting demand on electricity grids
- Cybersecurity and privacy in AI - Medical imaging diagnosis
- Artificial Intelligence and Cybersecurity Research



Fonte: Conferenza ENISA AI 7 giugno 2023

<https://www.euractiv.com/section/cybersecurity/news/eu-agency-maps-key-cybersecurity-issues-on-artificial-intelligence/>

Nel 2018 il Turing Award, una specie di Nobel dell'informatica, fu assegnato ai tre pionieri delle reti neurali, Yoshua Bengio, Geoffrey Hinton, Yann LeCun, per il loro contributo all'intelligenza artificiale.

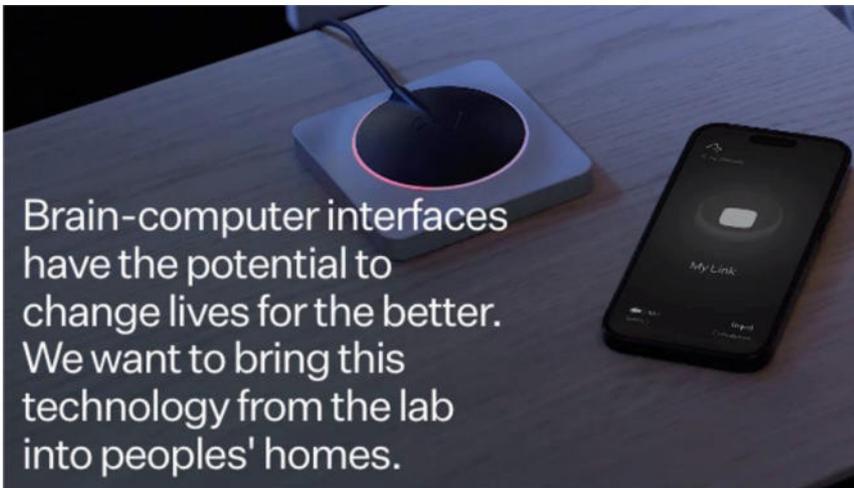
Il risultato spaziava dal riconoscimento di immagini a quello dei veicoli, dai dispositivi a guida autonoma ai robot della logistica, dagli assistenti digitali ai large language model come ChatGPT e Bard (Google).

Dopo il successo dell'IA generativa Yoshua Bengio è stato tra i firmatari della **lettera per una moratoria di sei mesi dei sistemi di intelligenza artificiale generativa**.

Geoffrey Hinton ha, invece, deciso nel mese di aprile 2023 di lasciare Google. Nell'intervista al New York Times, dichiara di aver deciso di dimettersi per intervenire sui rischi dell'IA senza danneggiare la reputazione di Google, che comunque ha agito finora in modo responsabile, a suo avviso. **Il successo di ChatGPT avrebbe attivato il meccanismo della concorrenza a causa del quale sarebbe ormai impossibile contenere lo sviluppo di sistemi pericolosi, a meno di una loro regolamentazione mondiale.**



Il 26 maggio FDA americana ha approvato la sperimentazione umana per i prodotti di neuralink che si occupa di interfacce neurali impiantabili



30 maggio 2023

Un appello lanciato dai leader delle più importanti realtà impegnate in questo ambito: nell'elenco figurano i vertici di OpenAI, Google DeepMind, Alphabet e decine di autorevoli rappresentanti del mondo accademico.

Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war.

Signatories:

AI Scientists Other Notable Figures

Geoffrey Hinton

Emeritus Professor of Computer Science, University of Toronto

Yoshua Bengio

Professor of Computer Science, U. Montreal / Mila

Demis Hassabis

CEO, Google DeepMind

Sam Altman

CEO, OpenAI

Dario Amodei

CEO, Anthropic

Dawn Song

Professor of Computer Science, UC Berkeley

Ya-Qin Zhang

Professor and Dean, AIR, Tsinghua University

Ilya Sutskever

Co-Founder and Chief Scientist, OpenAI

Shane Legg

Chief AGI Scientist and Co-Founder, Google DeepMind

Martin Hellman

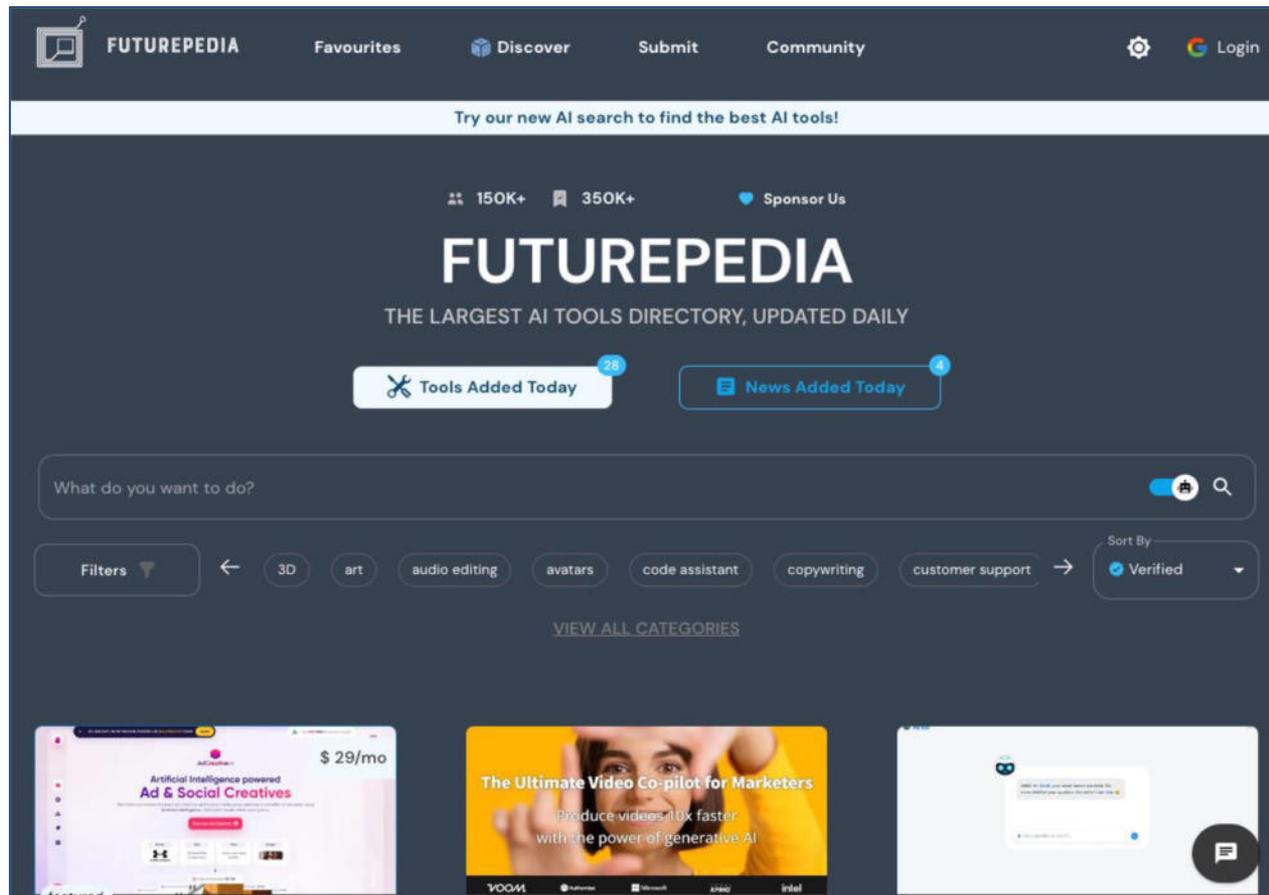
Professor Emeritus of Electrical Engineering, Stanford

James Manyika

SVP, Research, Technology & Society, Google-Alphabet

Fonte: <https://safe.ai/statement-on-ai-risk>

Molti nuovi tools online ogni giorno

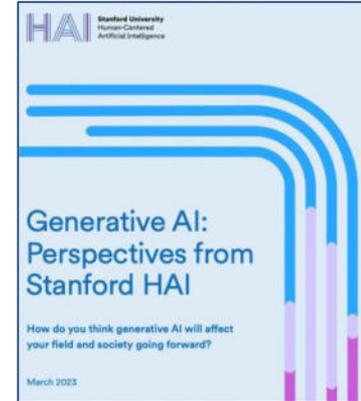


Fonte: hGps://www.futurepedia.io/

“La corrente ondata di intelligenza artificiale generativa è una sottocategoria dell'intelligenza artificiale che, basandosi su un input testuale, genera contenuti nuovi. **ChatGPT potrebbe scrivere un saggio, Midjourney potrebbe creare belle illustrazioni, mentre MusicLM potrebbe comporre un jingle.** La maggior parte delle moderne intelligenze artificiali generative si basa su **foundational model**, ovvero modelli di intelligenza artificiale addestrati su un ampio insieme di dati attraverso l'autoapprendimento a larga scala, e successivamente adattati a una vasta gamma di compiti specifici.

Le opportunità che questi modelli offrono alle nostre vite, alle nostre comunità e alla nostra società sono immense, così come i rischi che comportano. Da un lato, possono complementare senza soluzione di continuità il lavoro umano, rendendoci più produttivi e creativi, ma dall'altro potrebbero amplificare i pregiudizi che già sperimentiamo o minare la fiducia nelle informazioni...”

Fonte: https://hai.stanford.edu/sites/default/files/2023-03/Genera4ve_AI_HAI_Perspec4ves.pdf marzo 2023



Grazie per l'attenzione. Domande ?

