



AI e PAL

antonio memoli

11-12-13 marzo 2026
Sala Consilina
online webinar



Sequenza lineare 2/5



01

Introduzione AI

3 ore



02

AI e decisioni
di attuazione

3 ore

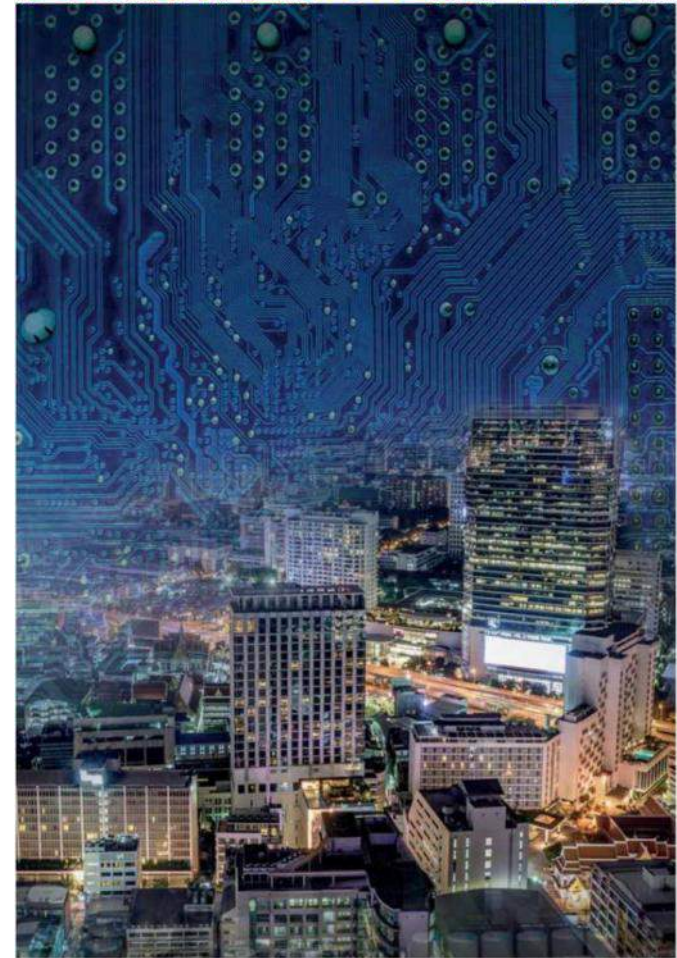


03

Test

Argomenti

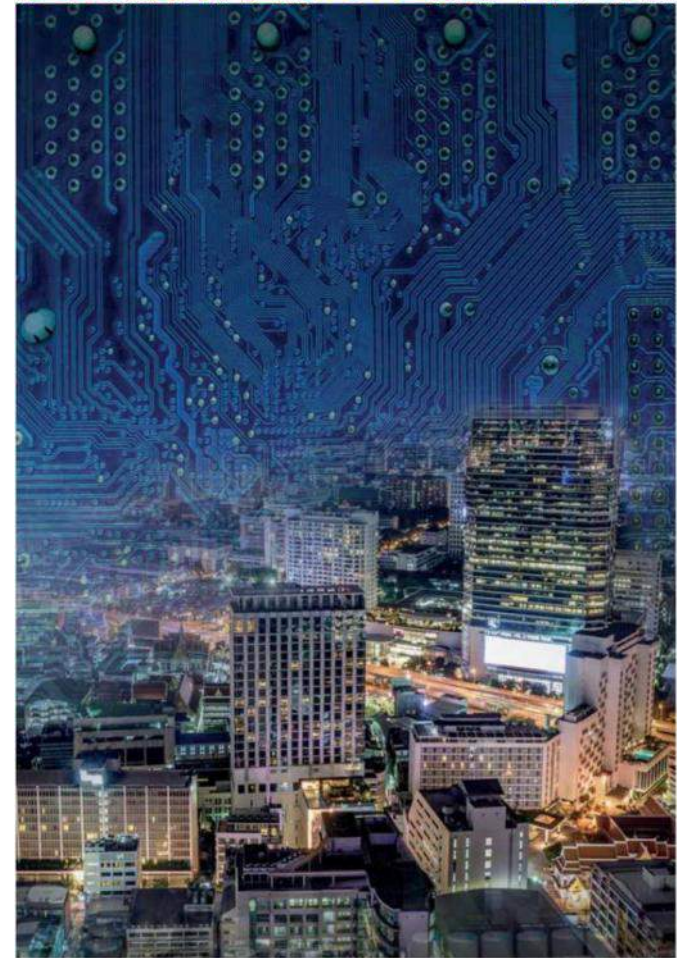
- Introduzione
- Strumenti di AI
- Cenni storici e terminologia
- Modelli di Intelligenza Artificiale
- Nuove evoluzioni dell'AI
- Normativa
- AI e Adottabilità dalla PA



Argomenti del giorno

- Introduzione
- Strumenti di AI
- Cenni storici e terminologia
- Modelli di Intelligenza Artificiale
- Nuove evoluzioni dell'AI
- Normativa

Ma prima di cominciare ...



Il settore delle telecomunicazioni sta cambiando

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

18 May · Comments

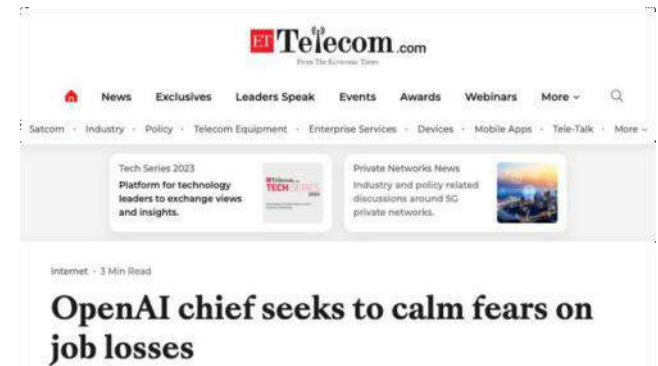


BT GROUP

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>

L'impatto a lungo termine è da chiarire

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/

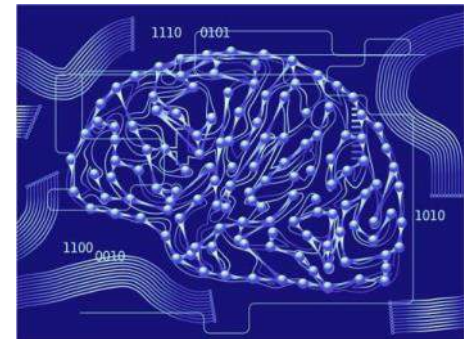
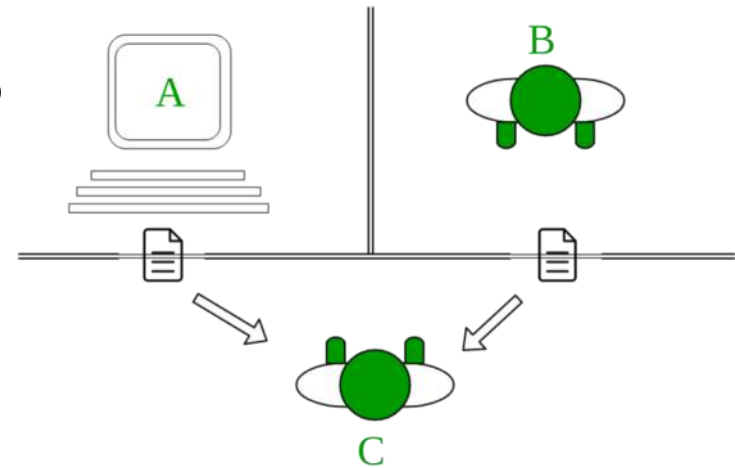
INTRODUZIONE

Intro - Definizione semplice di IA

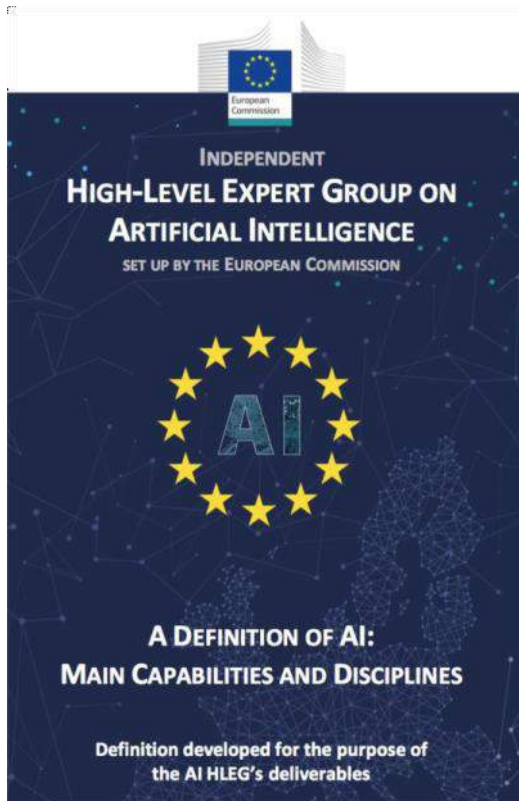
“L'intelligenza artificiale permette la progettazione di sistemi sia hardware sia software in grado di dotare le macchine di caratteristiche che vengono considerate tipicamente umane”.

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).



Intro - Definizione avanzata di IA



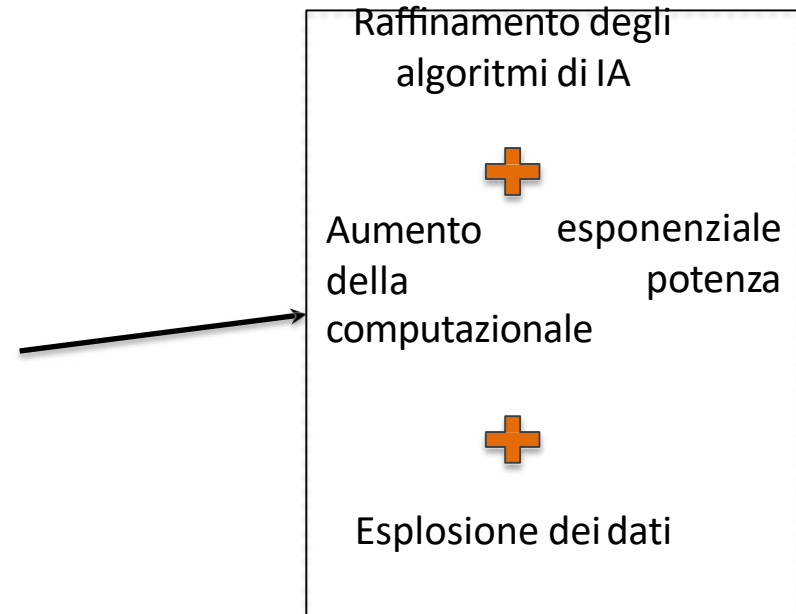
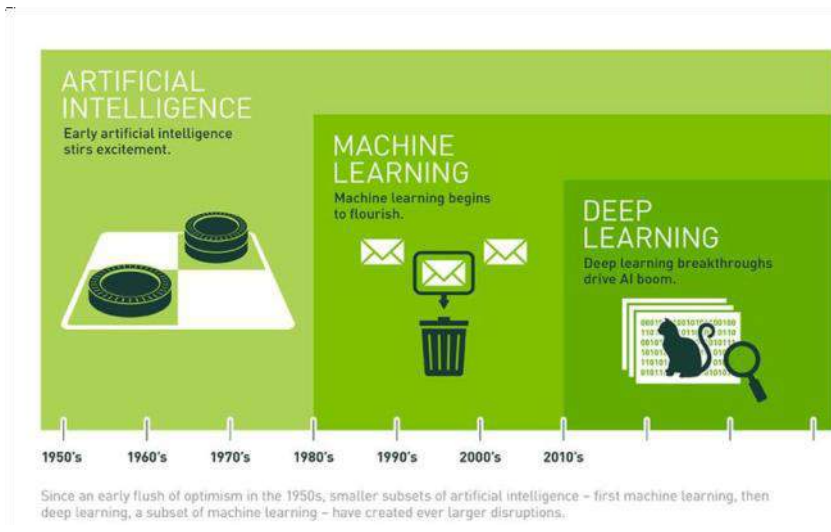
L'Intelligenza artificiale (IA) concerne i sistemi che manifestano un comportamento intelligente nell'analizzare il loro ambiente e nell'intraprendere azioni, con un certo grado di autonomia, per conseguire determinati obiettivi.

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

Introduzione

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.



Classificazione della 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 its 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

I servizi AI

Google

riconoscimento volti 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

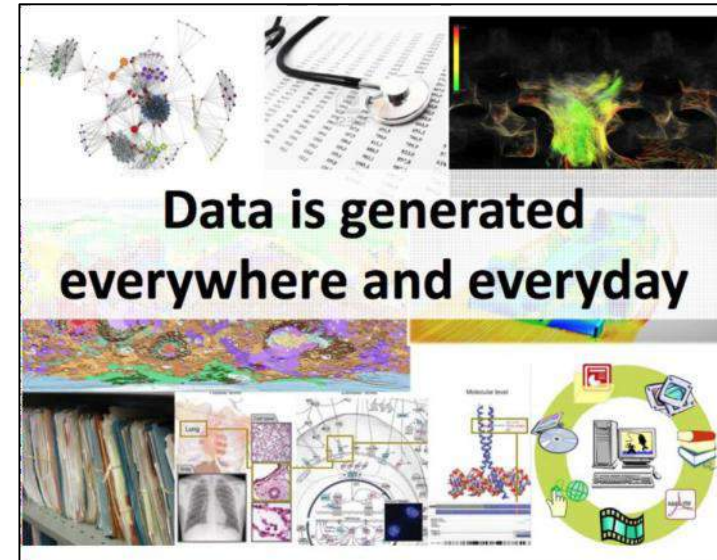
Raccomandazioni di prodog e gestione della rete di distribuzione...

Linkedin

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

E tantissimi altri servizi

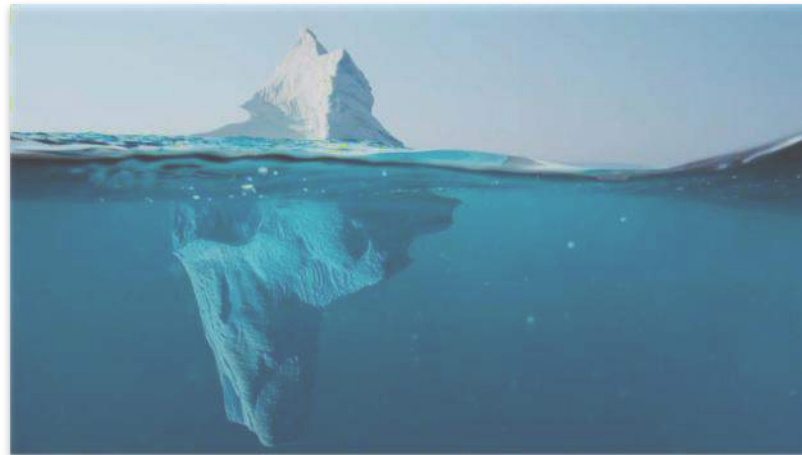
Ride-sharing apps, smart assistant, etc..



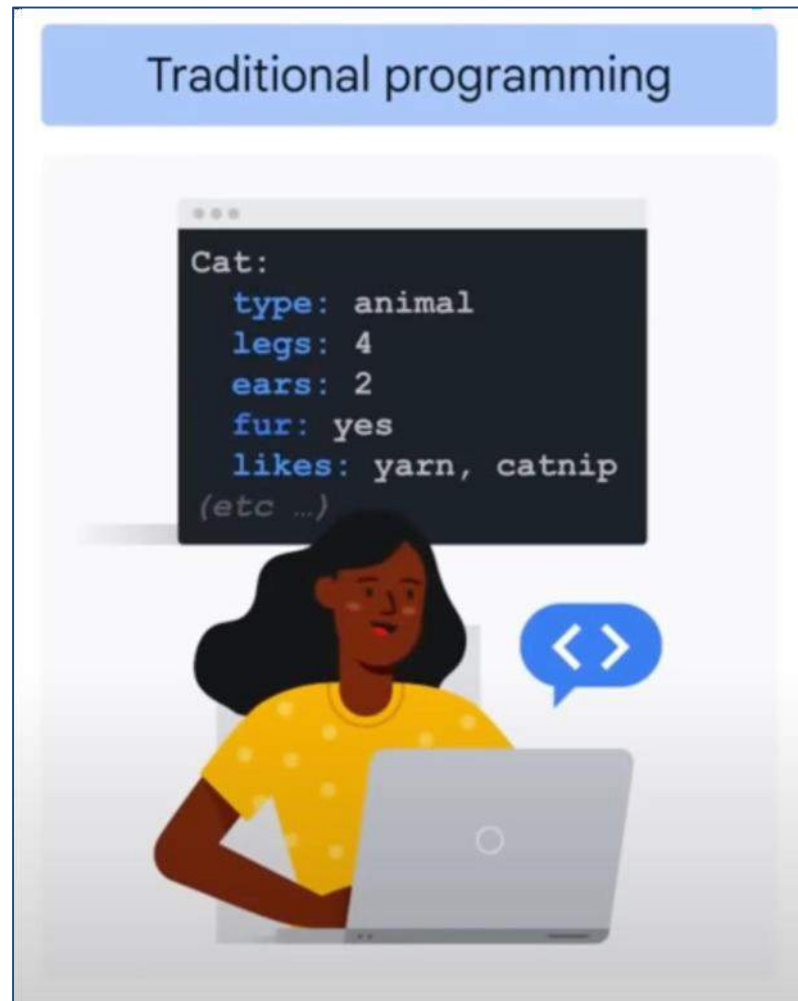
I servizi di AI di cui si parla

Modelli di linguaggio di grandi dimensioni (LLM) e la AI generativa

(ChatGPT, Bard,..)



Programmazione tradizionale



Fonte: Google Cloud Tech

Il ritorno delle reti neurali

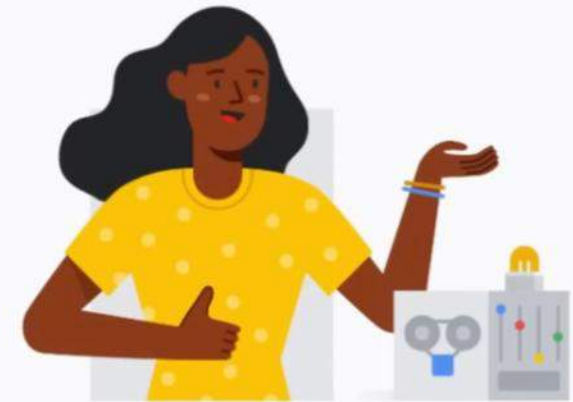
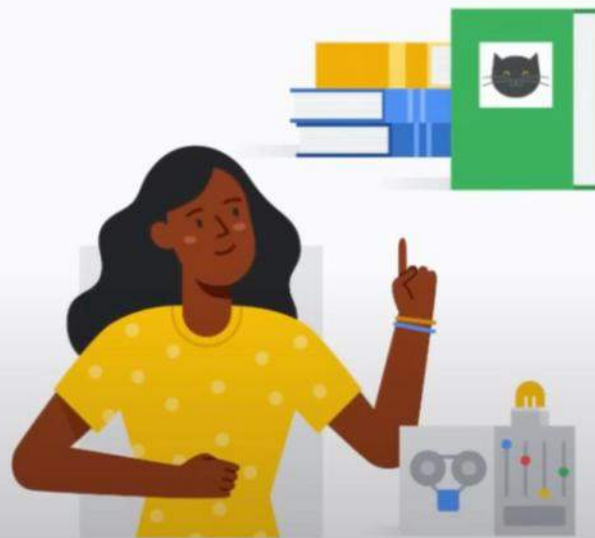


Fonte: Google Cloud Tech

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 ...

Dai modelli discriminativi a quelli generativi

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

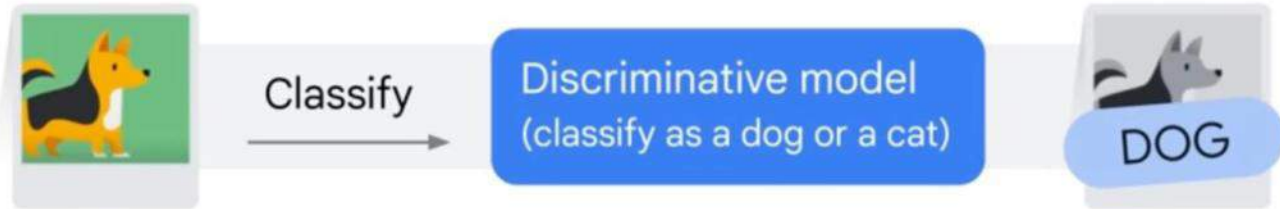


Generative

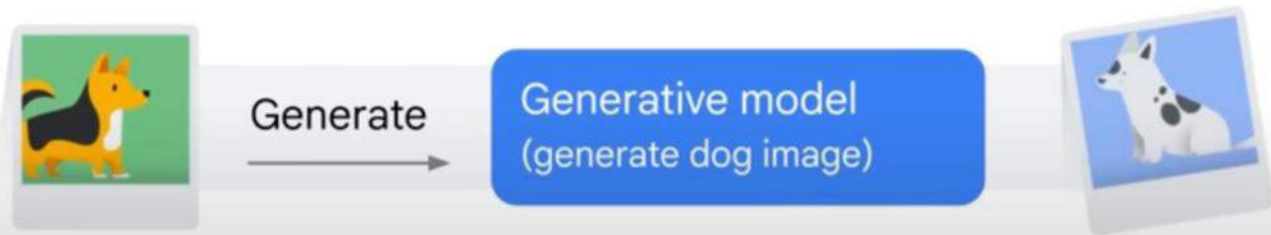
- Generates new data that is similar to data it was trained on
- Understands distribution of data and how likely a given example is
- Predict next word in a sequence

Dai modelli discriminativi a quelli generativi

Discriminative
technique



Generative
technique




ChatGPT

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/ 
Modifica dati su Wikidata · Manuale	

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

Attenzione

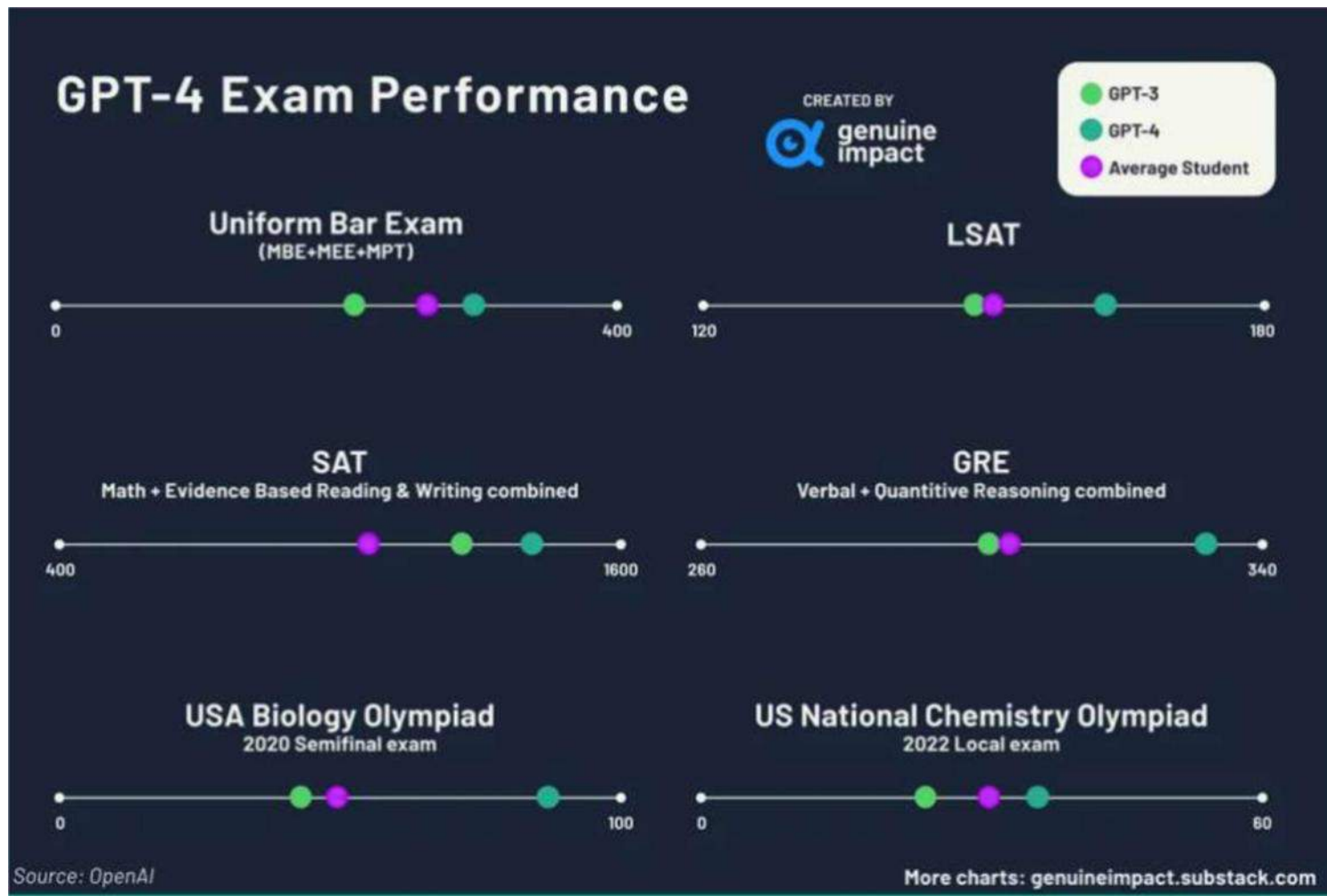
- Padronanza di linguaggio è un effetto collaterale di predizione delle parole di un testo che ha sorpreso anche gli sviluppatori. In sostanza, fa un ottimo lavoro nel prevedere ciò che un essere umano potrebbe scrivere
- Si basa su modelli statistici, non hanno la capacità di comprendere il testo che 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 è 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,

[hTps://www.ny4mes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html](https://www.ny4mes.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>

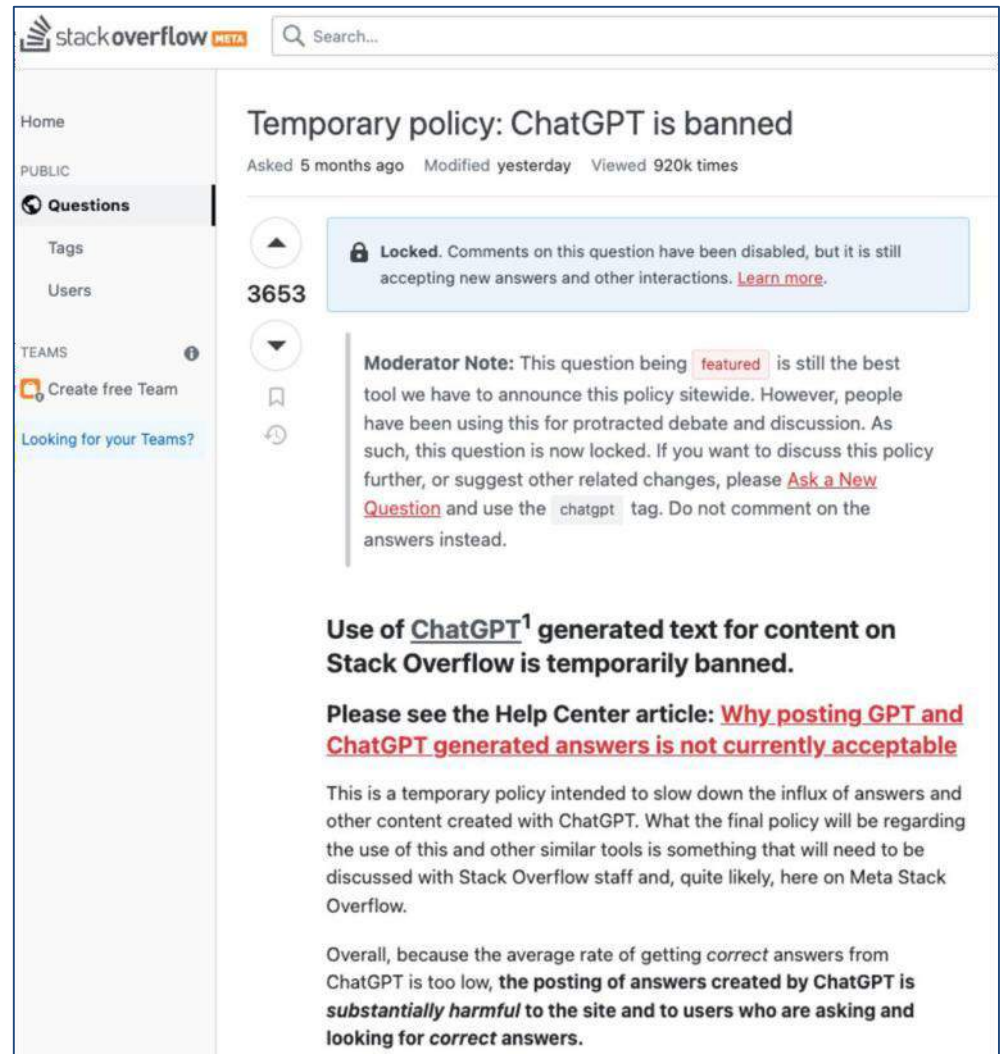
“Nel passaggio “evolutivo” della ChatGPT pare scomparsa la possibilità di richiudere la redazione di atti giudiziari veri e propri, a meno che tu non specifichi di essere un legale: in qualche modo la chatbot deve essere stata ben istruita sull’intervento necessario degli avvocati...”

Come usarlo ? (2) - Risolvere problemi tecnici

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 post. The title is "Temporary policy: ChatGPT is banned". It was asked 5 months ago, modified yesterday, and viewed 920k times. The post is locked, and comments are disabled. A moderator note explains that the question is locked because it is being used for protracted debate and discussion. The post also includes a section titled "Use of ChatGPT¹ generated text for content on Stack Overflow is temporarily banned." and a link to a Help Center article: "Why posting GPT and ChatGPT generated answers is not currently acceptable".

stackoverflow META Search...

Home

PUBLIC

Questions

Tags

Users

TEAMS

Create free Team

Looking for your Teams?

Temporary policy: ChatGPT is banned

Asked 5 months ago Modified yesterday Viewed 920k times

3653

Locked. Comments on this question have been disabled, but it is still accepting new answers and other interactions. [Learn more.](#)

Moderator Note: This question being **featured** is still the best tool we have to announce this policy sitewide. However, people have been using this for protracted debate and discussion. As such, this question is now locked. If you want to discuss this policy further, or suggest other related changes, please [Ask a New Question](#) and use the `chatgpt` tag. Do not comment on the answers instead.

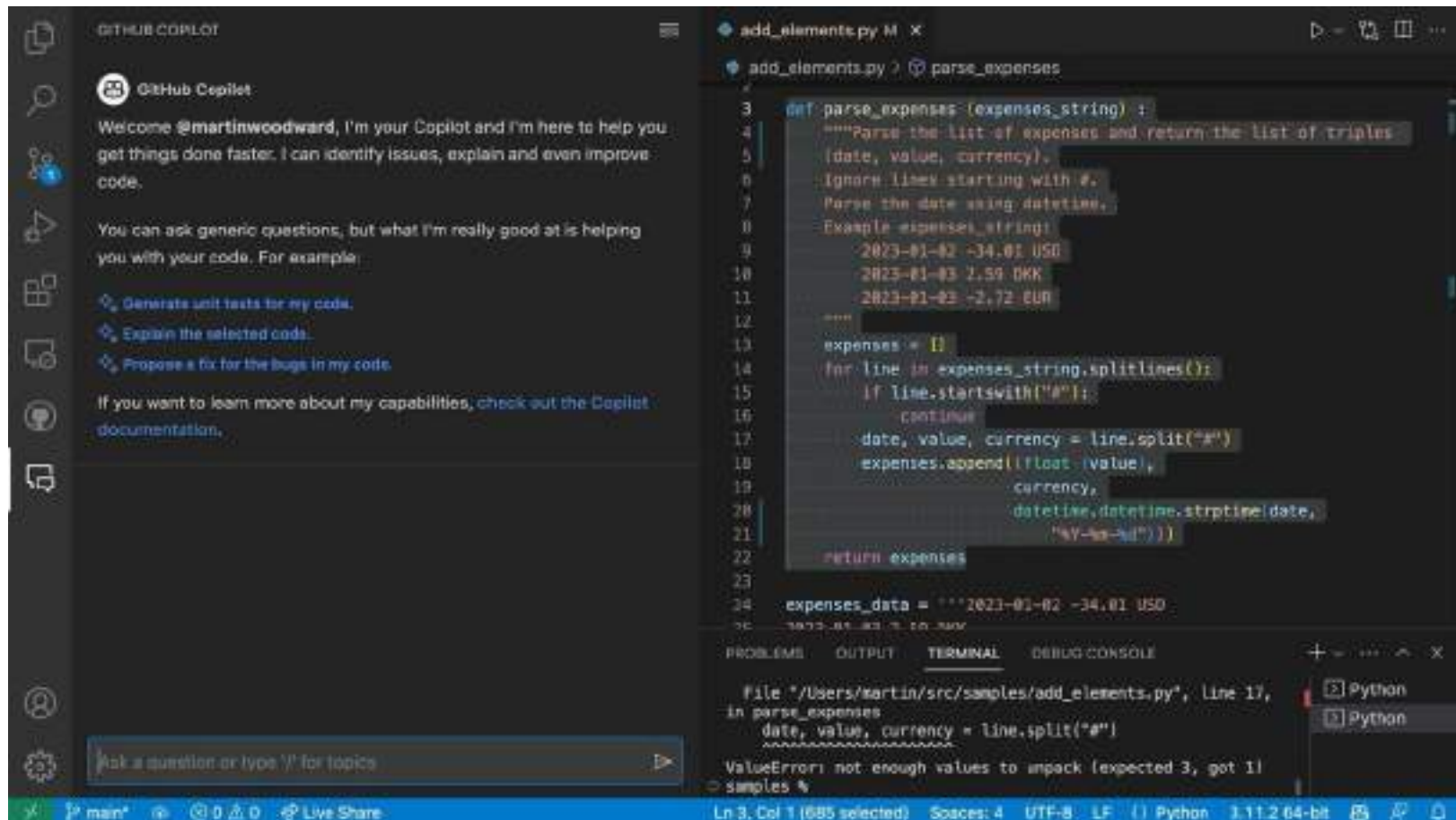
Use of ChatGPT¹ generated text for content on Stack Overflow is temporarily banned.

Please see the Help Center article: [Why posting GPT and ChatGPT generated answers is not currently acceptable](#)

This is a temporary policy intended to slow down the influx of answers and other content created with ChatGPT. What the final policy will be regarding the use of this and other similar tools is something that will need to be discussed with Stack Overflow staff and, quite likely, here on Meta Stack Overflow.

Overall, because the average rate of getting *correct* answers from ChatGPT is too low, **the posting of answers created by ChatGPT is *substantially harmful* to the site and to users who are asking and looking for *correct* answers.**

Come usarlo ? (3) - Supporto allo sviluppo



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 instructions on how to interact with Copilot. The main editor area 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 example expenses. At the bottom, the terminal window shows a 'ValueError: not enough values to unpack (expected 3, got 1)' error, which is highlighted in the code editor at line 17.

```
def parse_expenses(expenses_string):  
    """Parse the list of expenses and return the list of triples  
    (date, value, currency).  
    Ignore lines starting with #.  
    Parse the date using datetime.  
    Example expenses_string:  
    2023-01-02 -34.01 USD  
    2023-01-03 2.59 DKK  
    2023-01-03 -2.72 EUR  
    """  
    expenses = []  
    for line in expenses_string.splitlines():  
        if line.startswith("#"):  
            continue  
        date, value, currency = line.split("#")  
        expenses.append((float(value),  
                        datetime.datetime.strptime(date,  
                                                    "%Y-%m-%d")))  
    return expenses
```

```
expenses_data = '''2023-01-02 -34.01 USD  
2023-01-03 2.59 DKK
```

File */Users/martin/src/samples/add_elements.py', Line 17,
in parse_expenses
date, value, currency = line.split("#")
ValueError: not enough values to unpack (expected 3, got 1)
o samples %

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, 118120513530
               "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
```

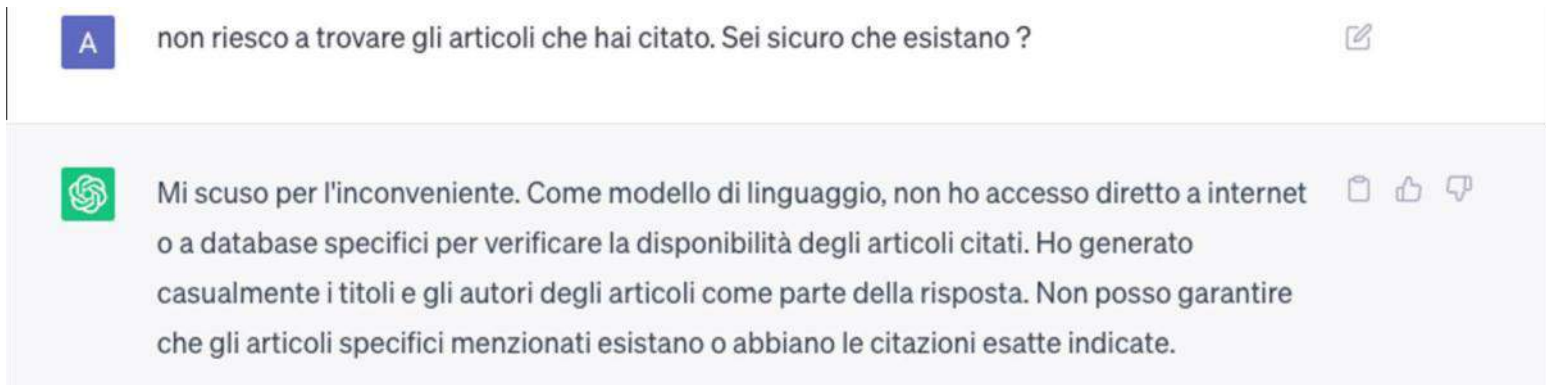
Fonte: <https://github.com/gventuri/pandas-ai>

Attenzione

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 ?



3 - Può essere soggetto ad allucinazioni

Domanda: "Quando Leonardo ha dipinto la Gioconda ?

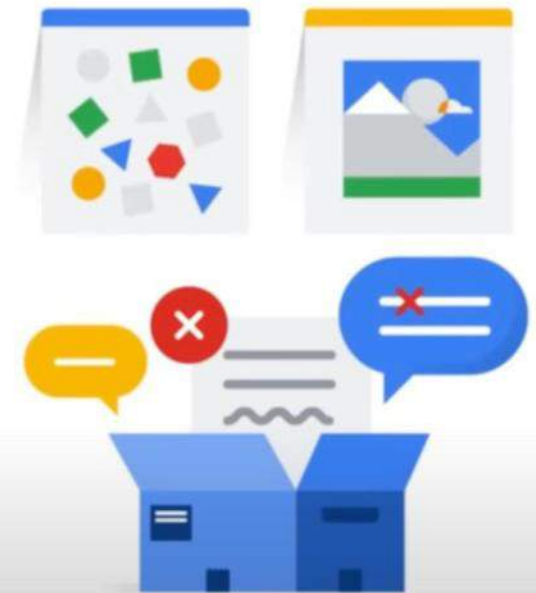
Risposta: "Leonardo da Vinci l'ha dipinta nel 1815 (risposta corretta 1503-1506)

Il modello soffre di allucinazioni

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



Non c'è solo ChatGPT

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

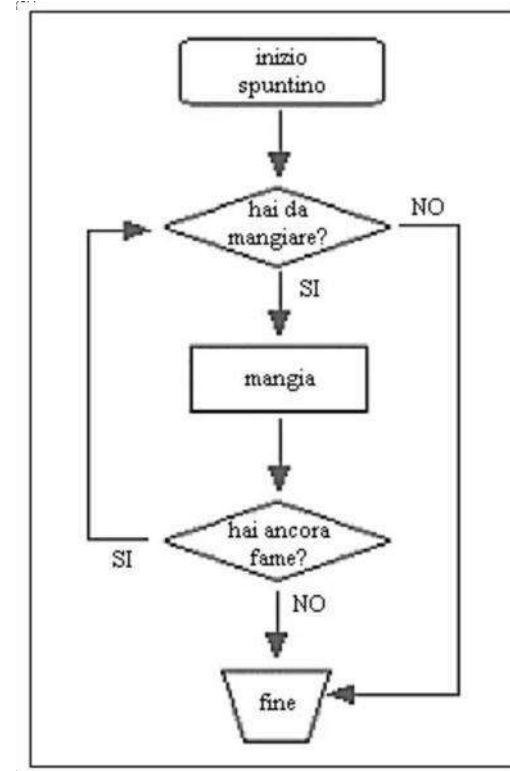
TERMINOLOGIA

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.

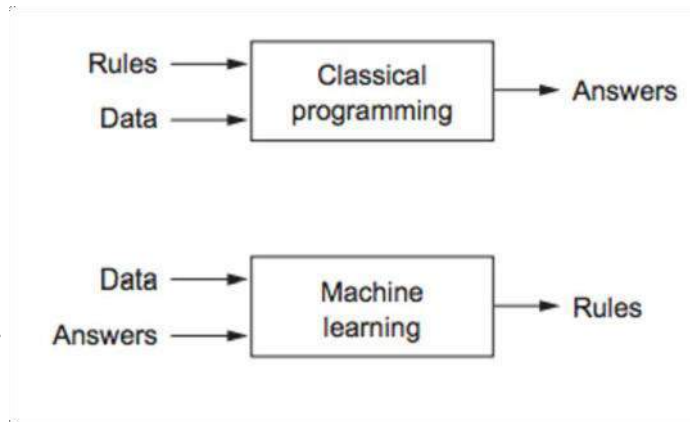


fonte: <https://www.informaticapertutti.com>

Programmazione vs Apprendimento (Learning)

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 statistica 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..

Terminologia (2)

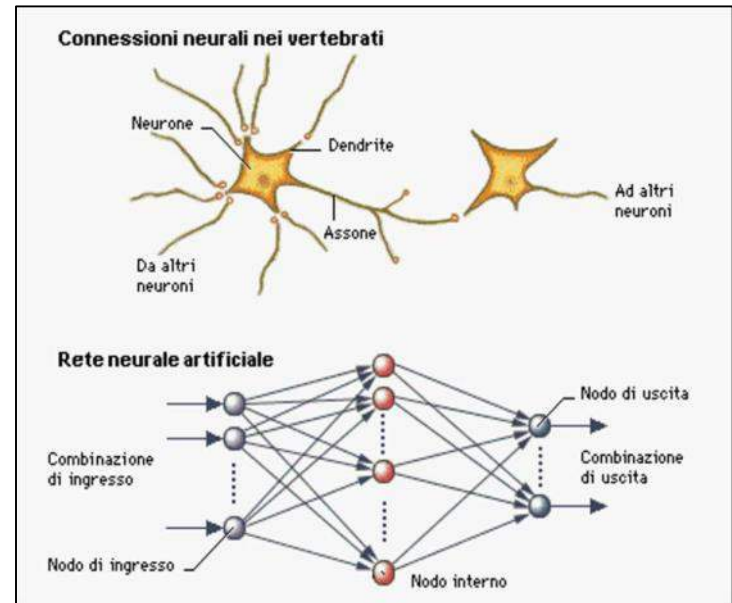
Reti neurali (Artificial Neural Networks)

Una rete neurale è un modello matematico composto di “neuroni” artificiali, che si ispira a una rete neurale biologica. Il modello ANN ha una certa somiglianza con gli assoni e dendriti in un sistema nervoso.

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

La rete neurale è composta da “unità” chiamate neuroni, organizzati in strati successivi. Ciascun neurone è tipicamente collegato ai neuroni dello strato successivo tramite connessioni pesate.

Il primo modello di rete neurali fu proposto nel 1943 da McCulloch e Pitts nei termini di un modello computazionale dell'attività nervosa. A questo modello sono seguiti altri proposti da John von Neumann, Marvin Minsky, Frank Rosenblatt, e molti altri.



Terminologia (3)

Machine learning

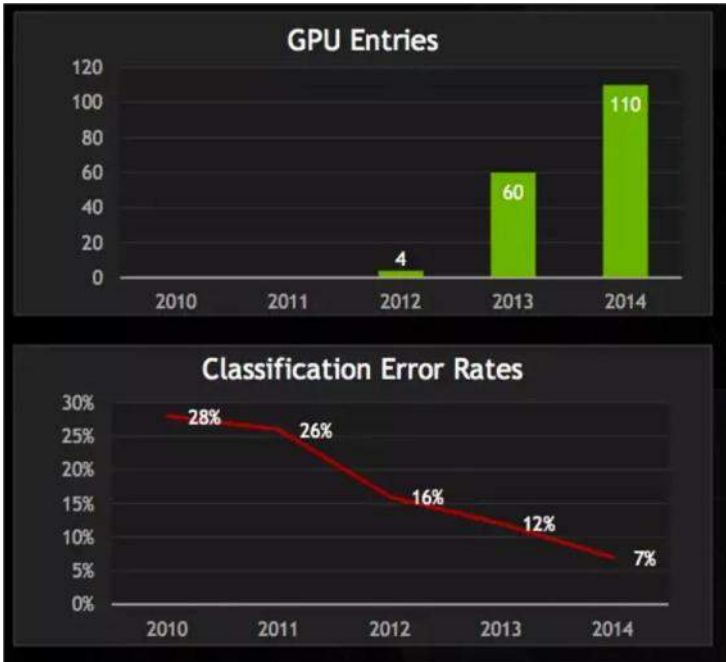
Con Machine Learning si parla di meccanismi che permettono a una macchina di migliorare le proprie capacità e prestazioni nel tempo. La macchina sarà in grado di imparare a svolgere determinati compiti migliorando, tramite l'esperienza, le proprie capacità, le proprie risposte e funzioni. Si basa su: 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 scientifico 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

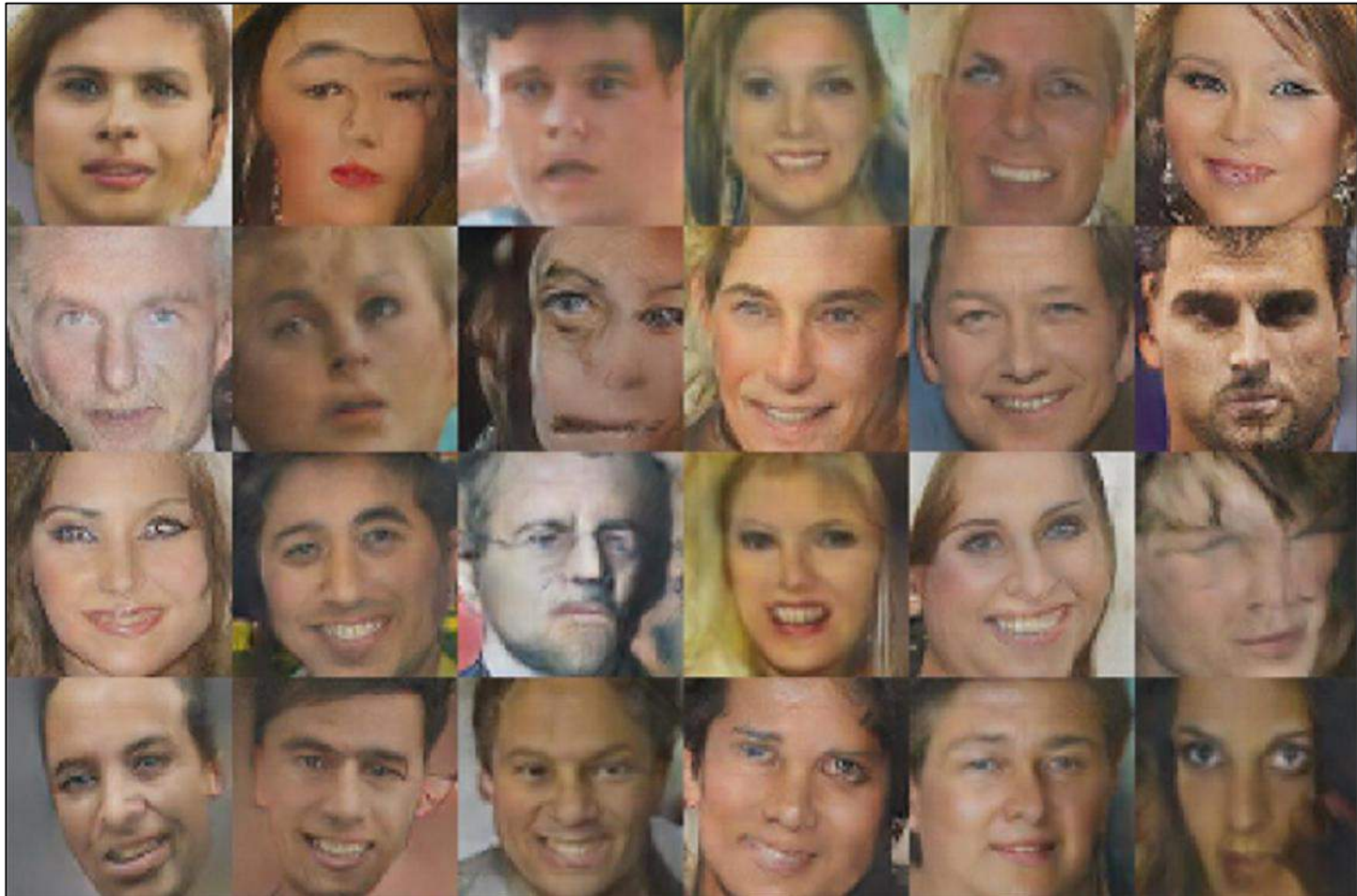
L'intelligenza artificiale generativa o AI generativa è un tipo di sistema di intelligenza artificiale (AI) in grado di generare testo, immagini o altri media in risposta a richieste.

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 linguaggi di base GPT-5 e Bard, un chatbot creato da Google utilizzando il loro modello di base LaMDA.

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

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 (creato nel 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

Immagini generate dalla IA - oggi

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

Immagini generate dalla IA - oggi

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

Video generati dalle IA

Deepfakes (deep learning + fakes) video generati dalla IA.

Es: Un finto video di Obama che critica Trump

Si corre il rischio di entrare in un'epoca in cui non sarà più possibile determinare se il contenuto di un oggetto multimediale 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>

Esempio: BBC

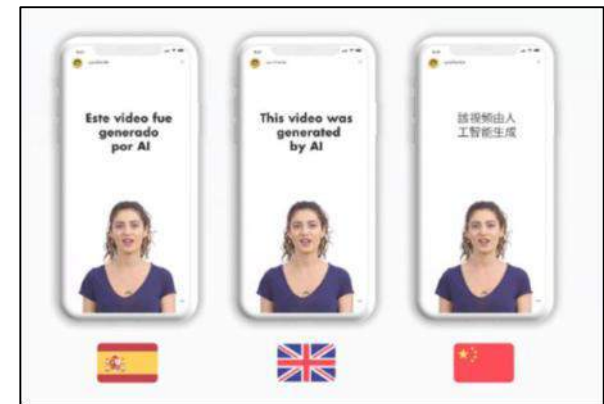
Il giornalista della BBC Matthew Amroliwala parla solo inglese, ma usando un software di IA sembra che parli spagnolo, mandarino e hindi. La tecnica utilizza un software che sostituisce una faccia originale con una faccia generata in cui vengono modificate le espressioni e sincronizzate ad un flusso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



Nuovi rischi - Vishing

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.



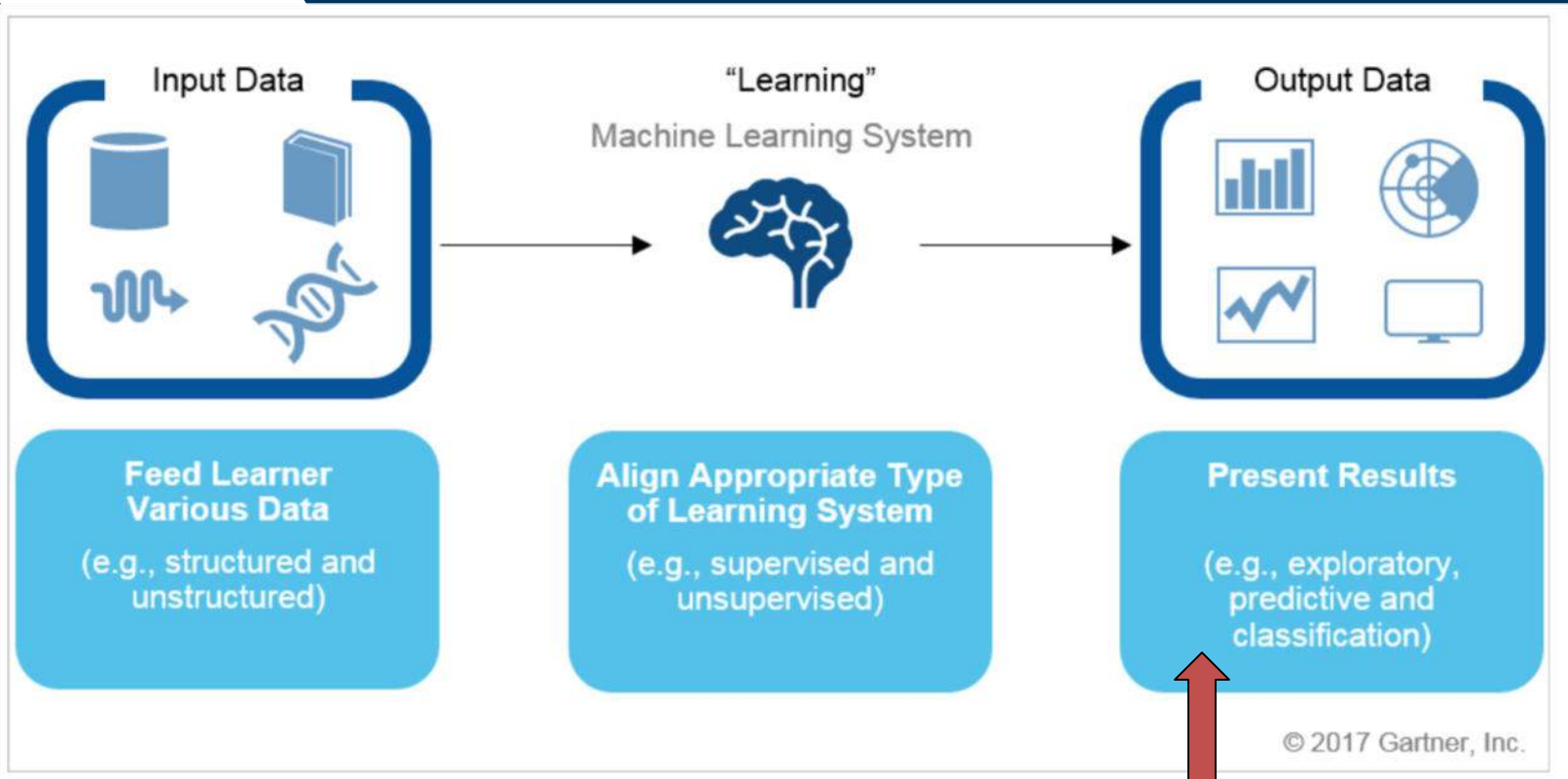
AI Could Escalate New Type Of Voice Phishing Cyber Attacks

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

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

STRUTTURA DI APPLICAZIONI DI IA

IA - Schematizzazione di un flusso di processo



NB: Nel 2017 ancora non si parla di output generativi

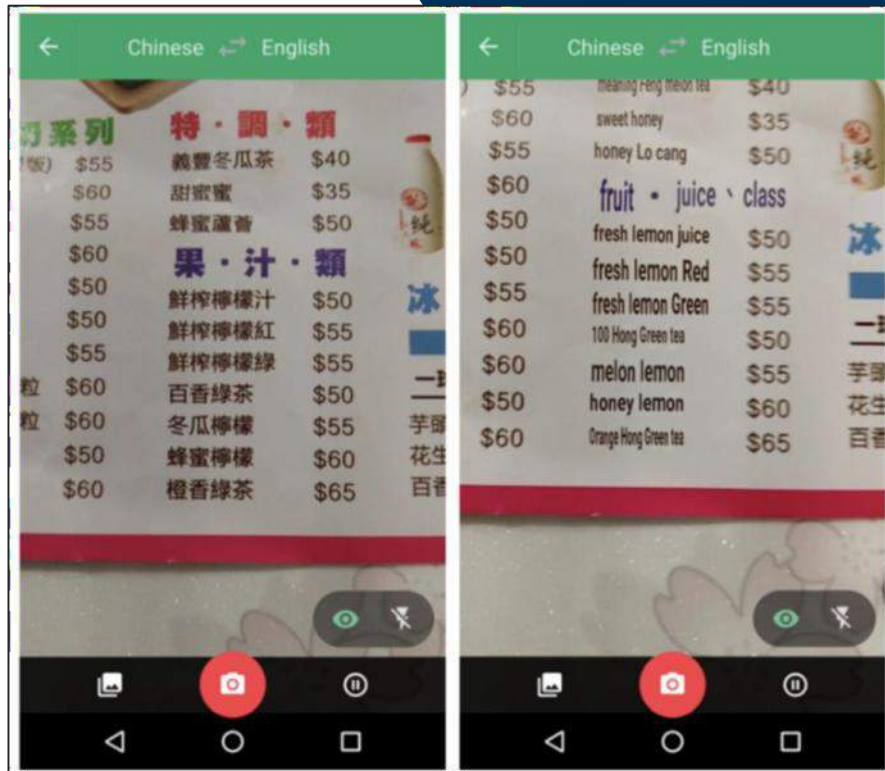
IA – Learning (Imparare dai dati)

1Dati 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.

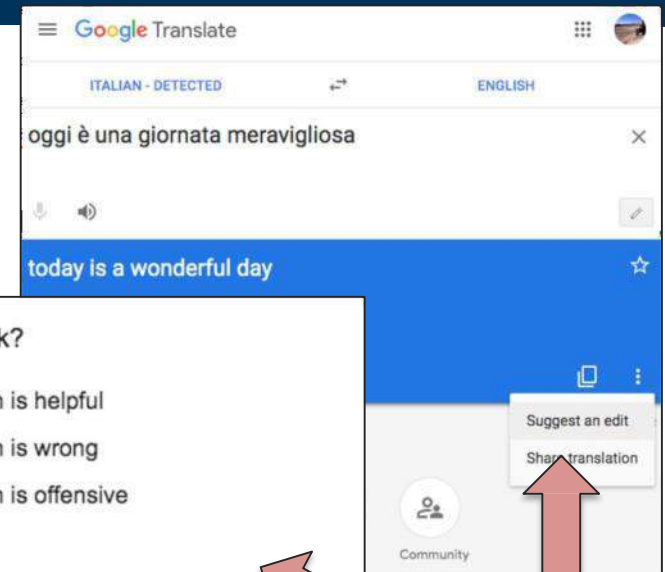
2Esempi di output atteso: in un'attività di riconoscimento vocale, queste potrebbero essere trascrizioni generate dall'uomo di file audio. In un'attività di riconoscimento delle immagini, i risultati previsti potrebbero essere etichette (tag) come "cane", "gatto" e così via.

3Un 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](#).

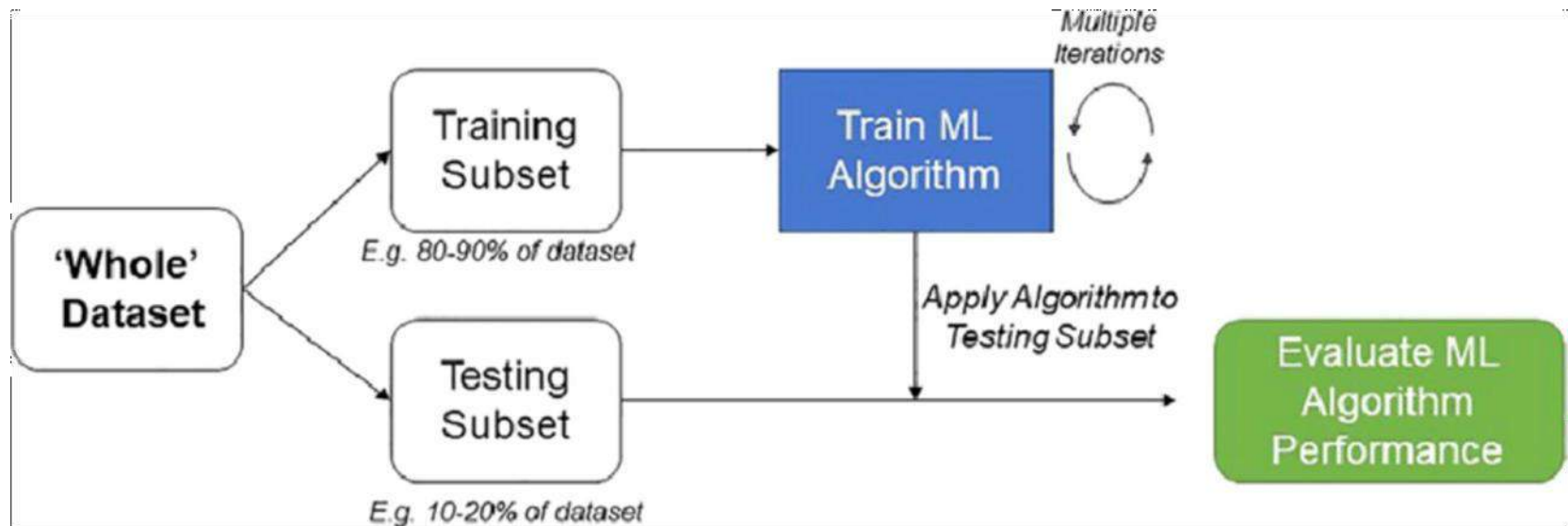
CANCEL SEND

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>

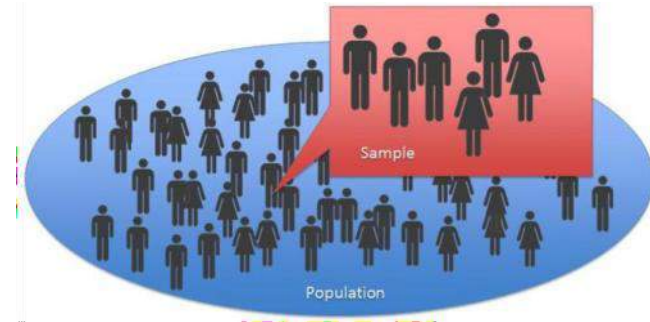
IA – Il processo del machine learning

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.



Bias - pregiudizio

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.

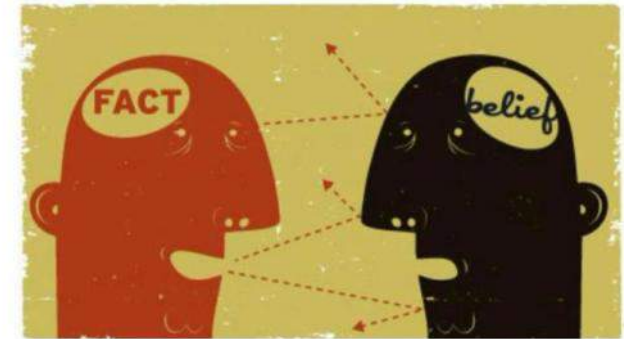
Qualche esempio di bias

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..).

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>

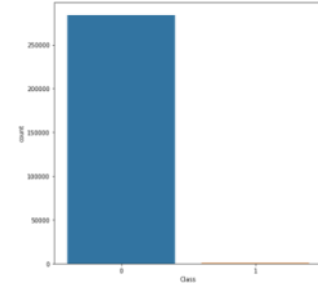


Rischio Bias per:

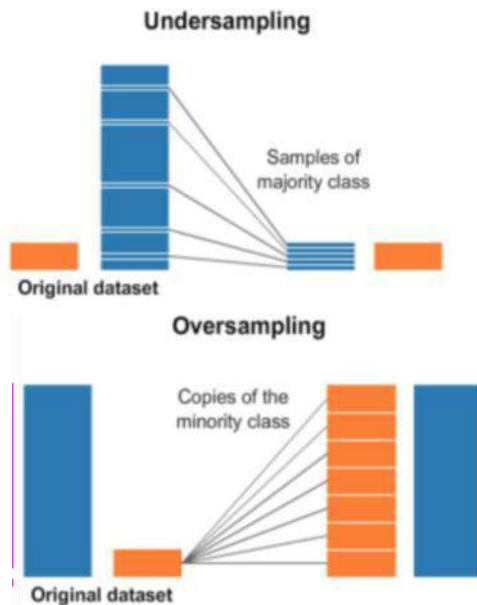
- Etnia
- Sesso
- Nazionalità
- Religione
- Orientamento sessuale
- Età
- Etc...

Attenzione al bilanciamento delle classi dei dati

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

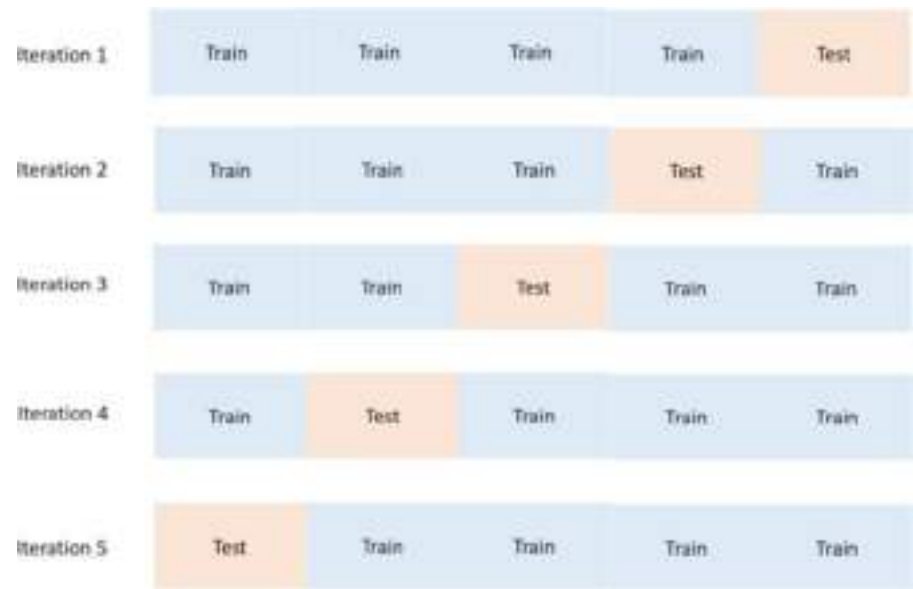
Sovracampionamento è il processo di generazione di dati aggiungendo copie delle istanze della classe sottorappresentata.



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

Valutare un modello – la cross validation

- 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.



IA – Tipologie di apprendimento

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.

TYPES OF MACHINE LEARNING

Supervised Learning



Train an algorithm on a labeled data set to predict the correct output value for unseen inputs.

- ✓ Input / Output
- ✓ Labeled data
- ✓ 'Replicate the right answers'
- ✓ Classification, prediction
- ✓ Image recognition, email spam filtering, forecasting

Unsupervised Learning



Train an algorithm to find similarities or abnormalities in a data set.

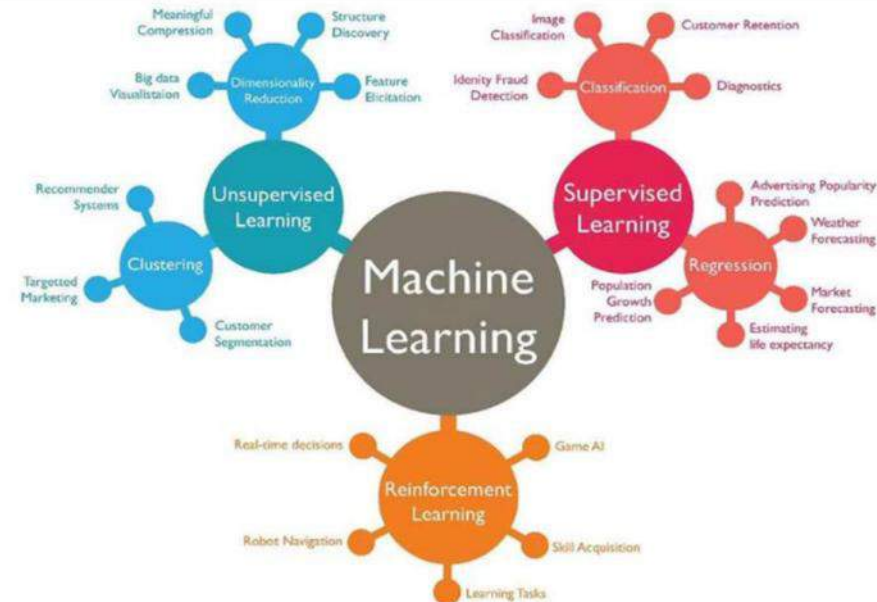
- ✓ Input
- ✓ Unlabeled data
- ✓ 'Find patterns in data'
- ✓ Clustering, association
- ✓ Anomaly detection, customer segmentation

Reinforcement Learning



Learn through trial and error from interaction with an environment.

- ✓ States & actions
- ✓ No data set
- ✓ 'Find actions that maximize reward'
- ✓ Decision making
- ✓ Learning to play a game, movie recommendation system



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



Esempio di supervised learning - Previsione del reddito

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

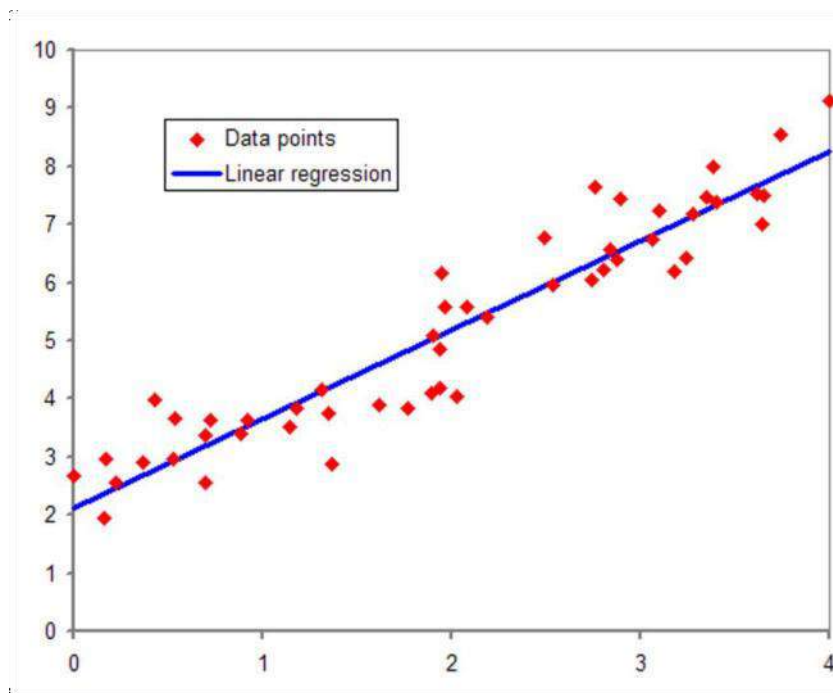
Esempio reale: Artificial intelligence diagnoses lung cancer

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

Algoritmi di supervised learning - Regressione lineare

Regressione lineare

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



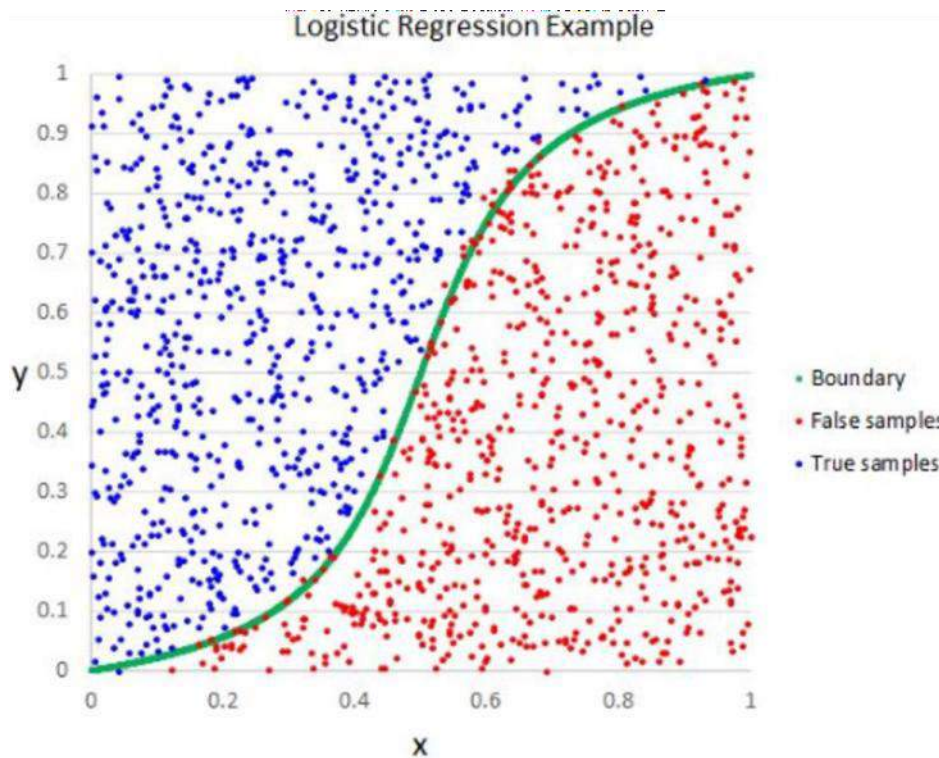
Caso d'uso

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

Algoritmi di supervised learning - Regressione logistica

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.

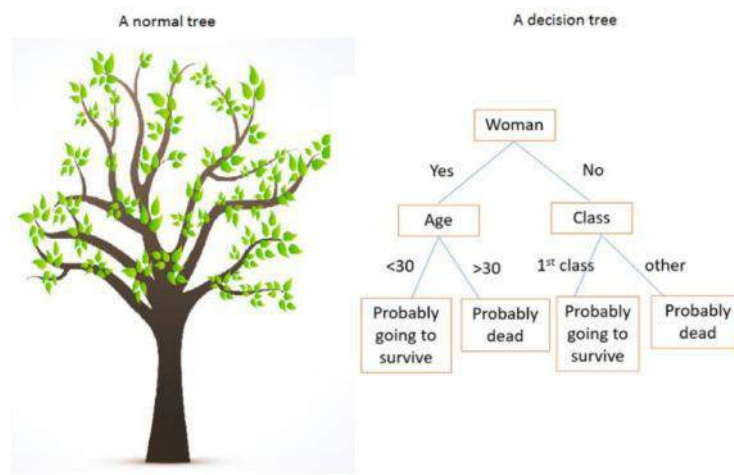
Algoritmi di supervised learning – Alberi decisionali

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>

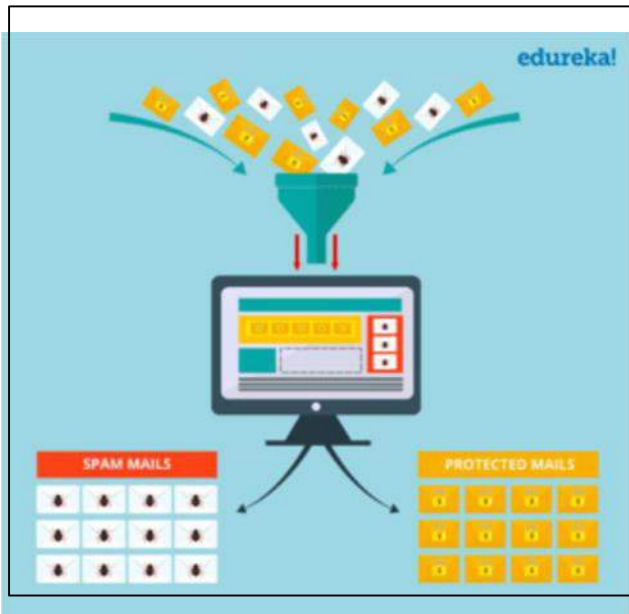
Algoritmi di supervised learning – Naive Bayes

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 tratti di spam è alta).

Caso d'uso

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



Unsupervised learning

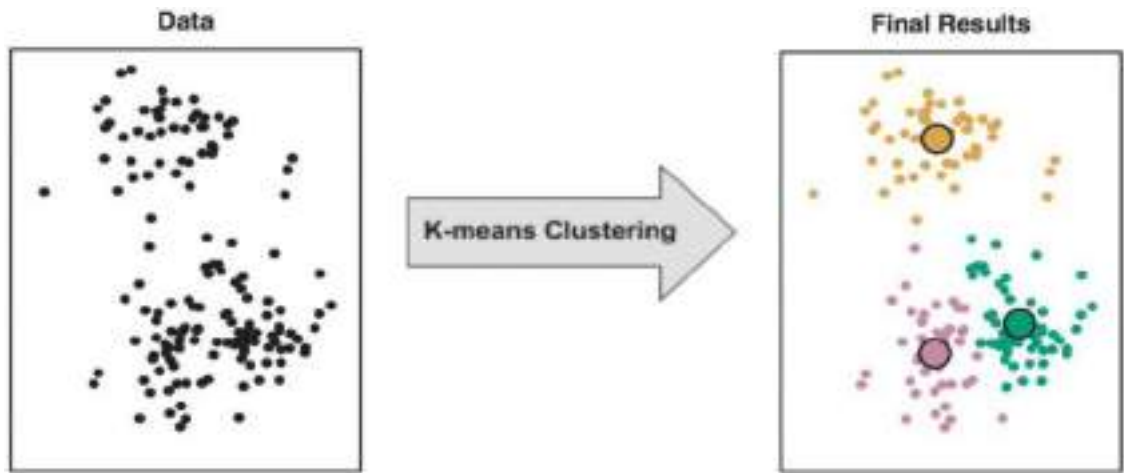
Che cos'è

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

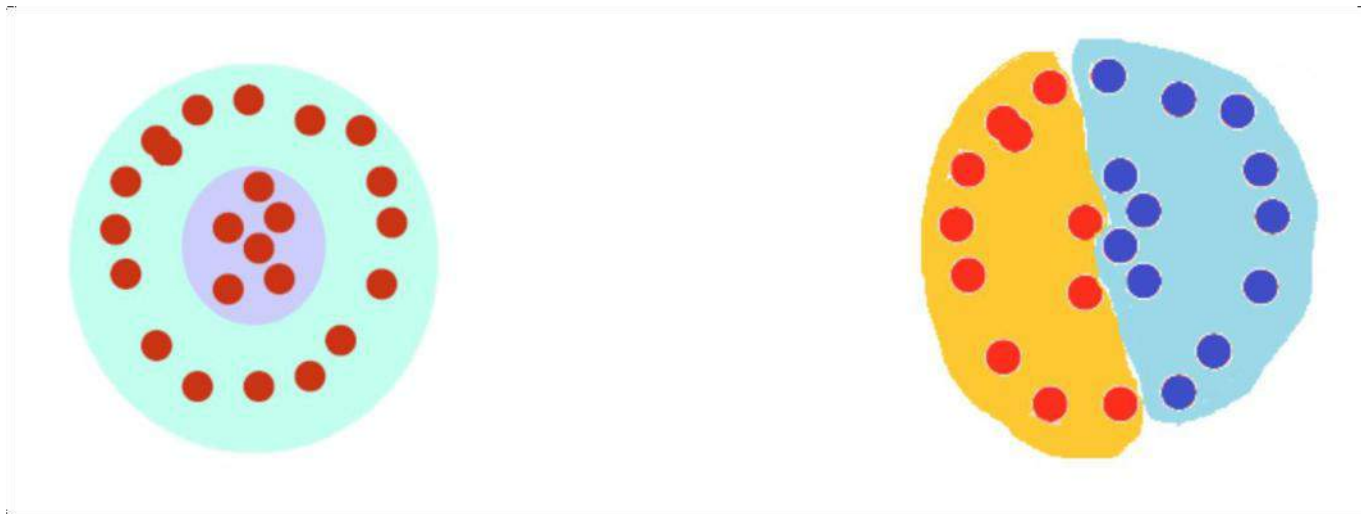
Non è chiaro come classificare i dati e si vuole che l'algoritmo trovi i modelli per classificare i dati in modo autonomo.

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 di altri cluster.



Unsupervised learning

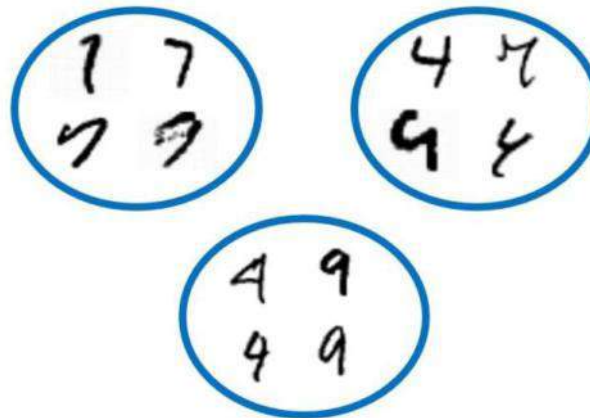
Come funziona

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

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

Raggruppare tra loro numeri scritti a mano

Esempio



Dataset: MNIST handwritten digits dataset

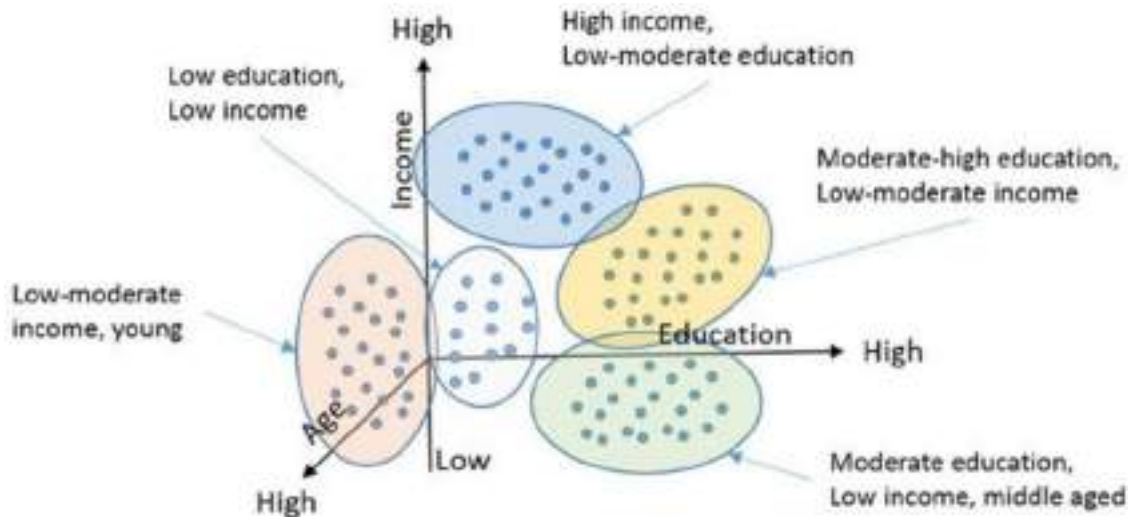
Algoritmi di unsupervised learning – K-means clustering

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

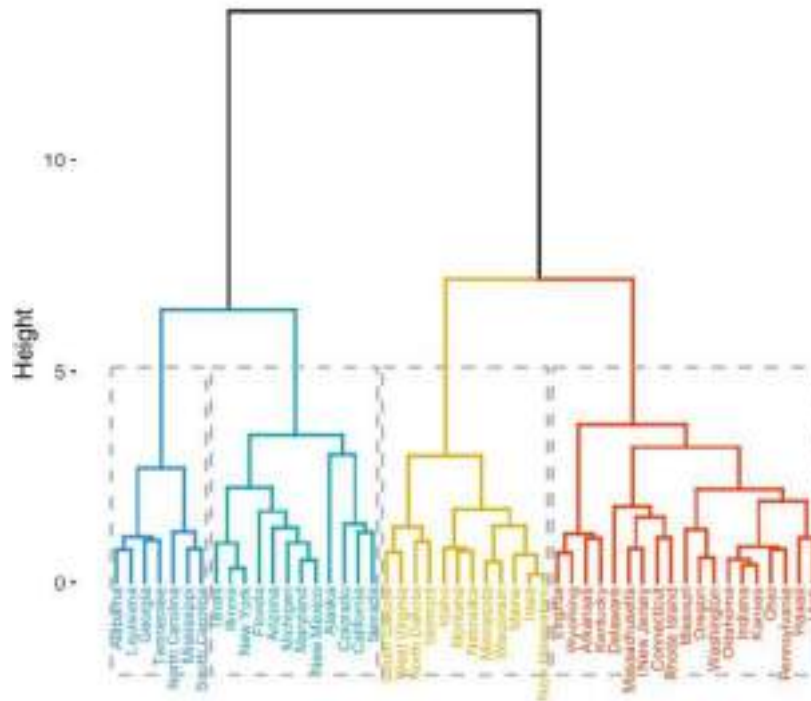
- Raggruppare i clienti delle carte fedeltà in gruppi segmentati



Algoritmi di unsupervised learning – Hierarchical clustering

Hierarchical clustering

L'algoritmo raggruppa o divide i cluster secondo una gerarchia ad albero per formare un sistema di classificazione.



Caso d'uso

- Segmentare i clienti in gruppi di caratteristiche distinte per meglio organizzare campagne di marketing.

- La radice rappresenta l'intero set di dati - Una foglia rappresenta un singolo oggetto nel set di dati;
- Un nodo interno rappresenta l'unione di tutti gli oggetti nella sua sottostruttura;
- L'altezza di un nodo interno rappresenta la distanza tra il nodo e i suoi nodi figli.

IA per la classificazione dei cittadini/consumatori

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

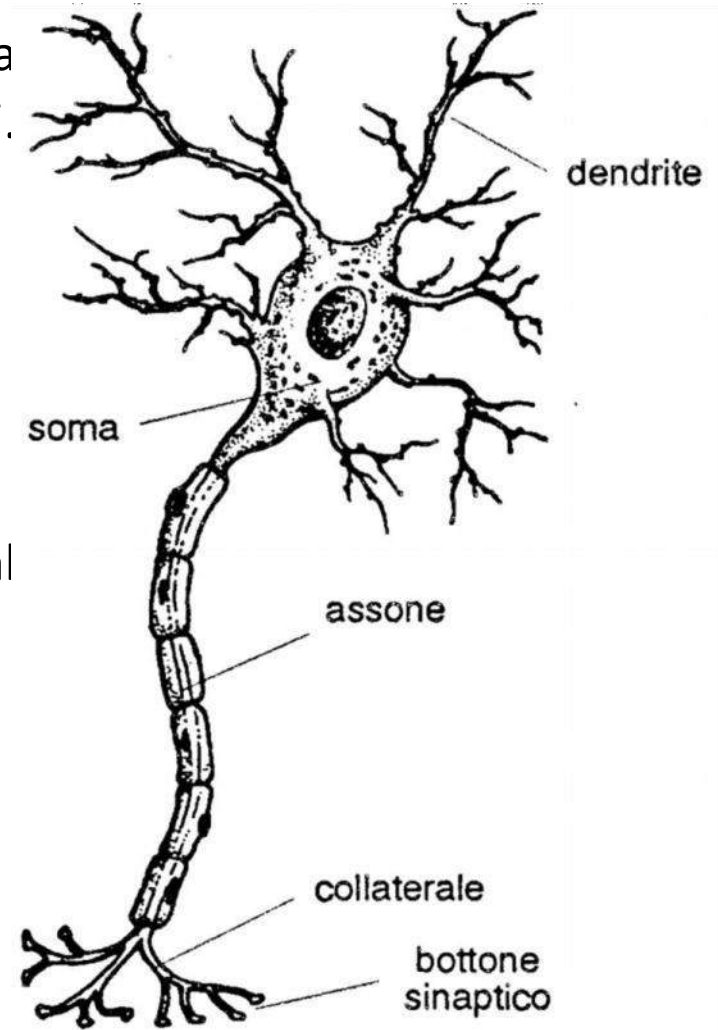
Sistema nervoso umano

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

Un neurone è caratterizzato da:

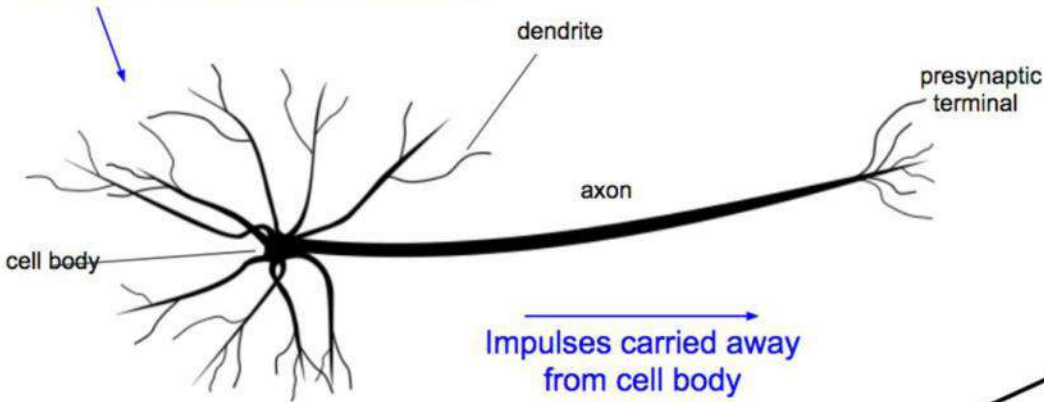
- corpo cellulare: l'unità di calcolo;
- assone: linea di trasmissione in uscita;
- dendriti: le zone ricettive.

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



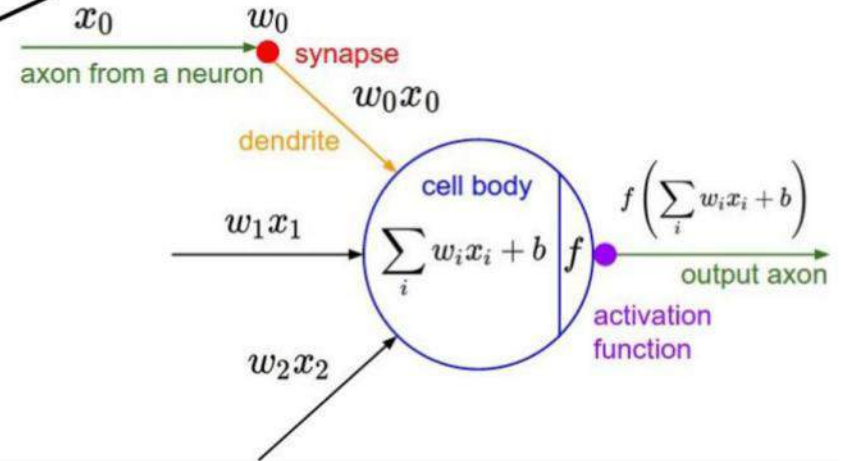
Neurone artificiale

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.

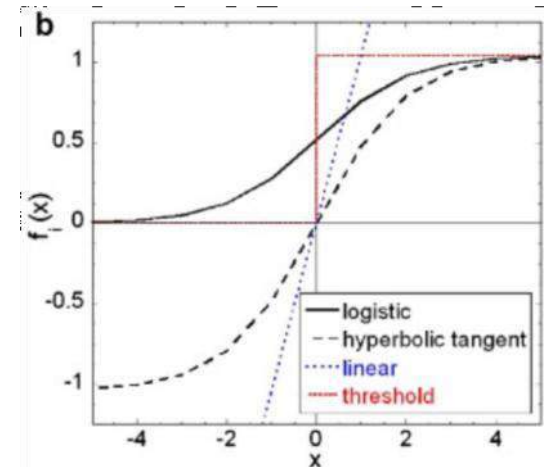
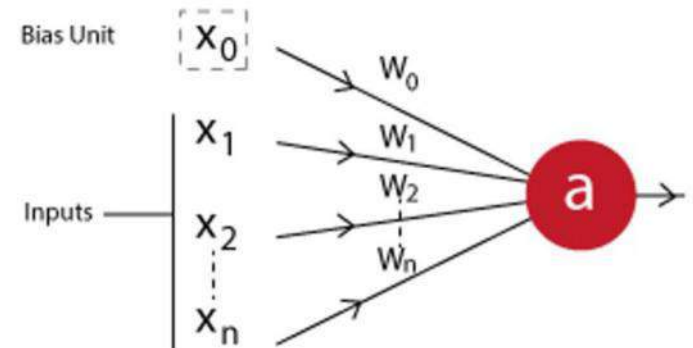
a : **Output** del neurone che è calcolata come:

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

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.

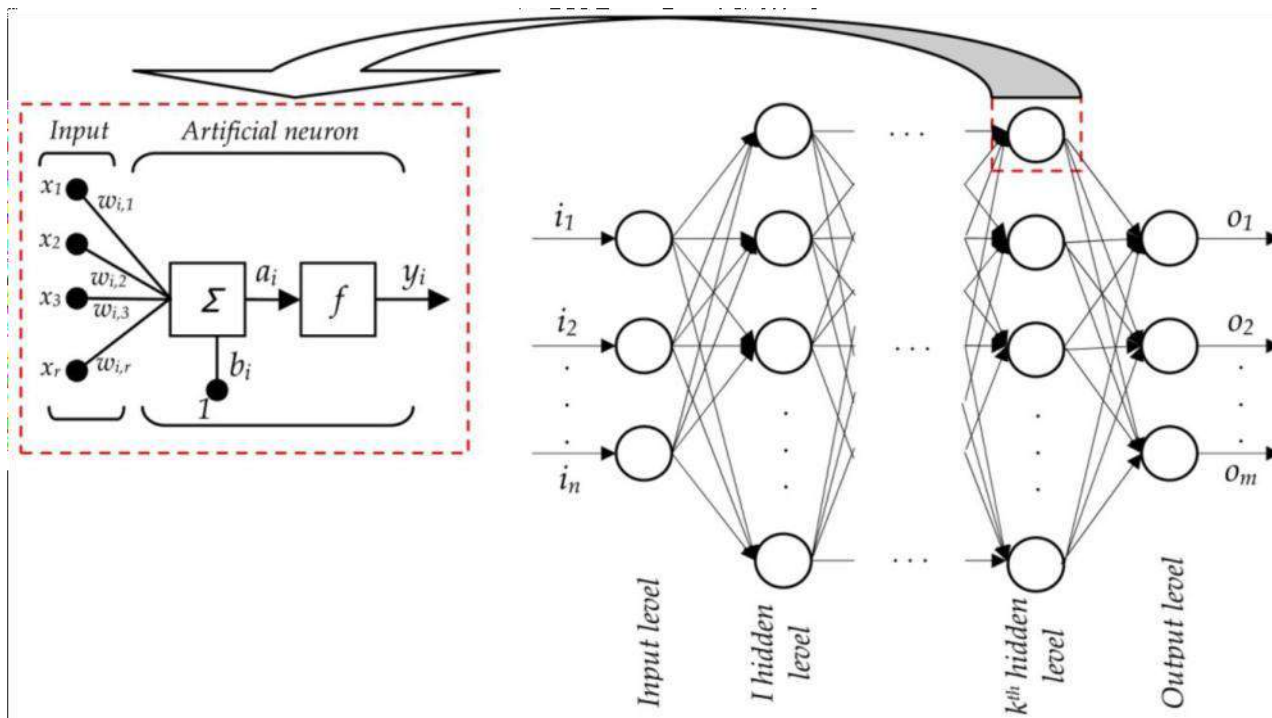
Diagram 1: Single NN Working



Rete neurale

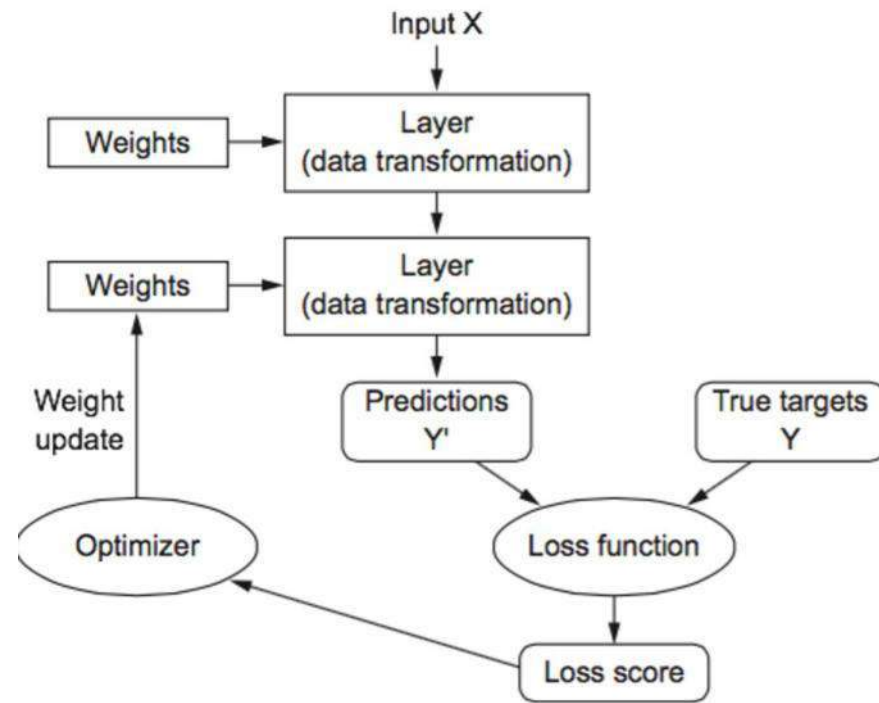
Una rete neurale è costituita da:

- Un insieme di nodi (neuroni), o unità connesse da collegamenti.
- Un insieme di pesi associati ai collegamenti.
- Un insieme di soglie o livelli di attivazione.



Processo di apprendimento

- 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

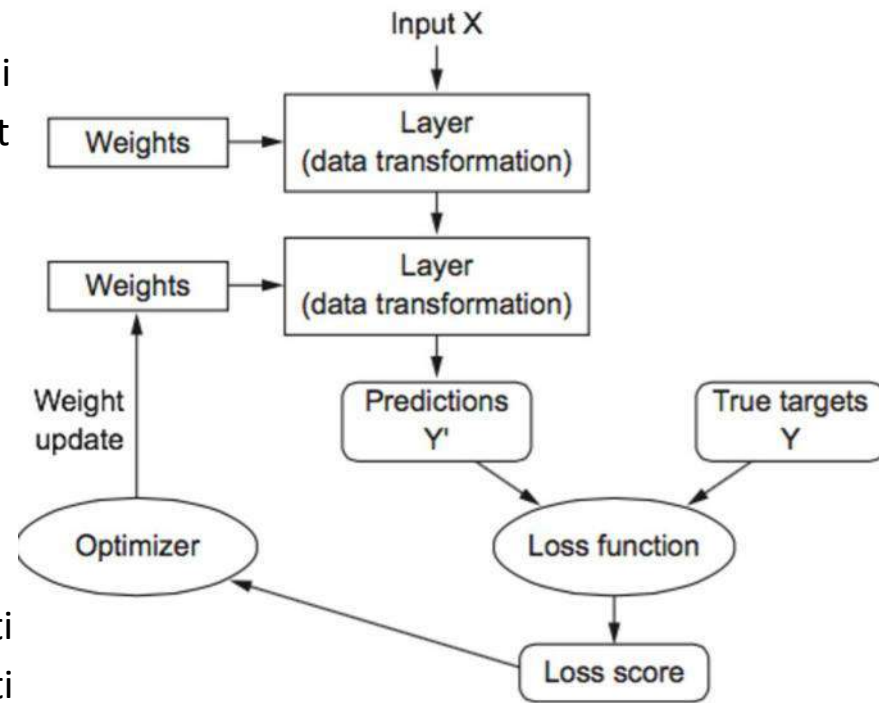


Processo di apprendimento

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.

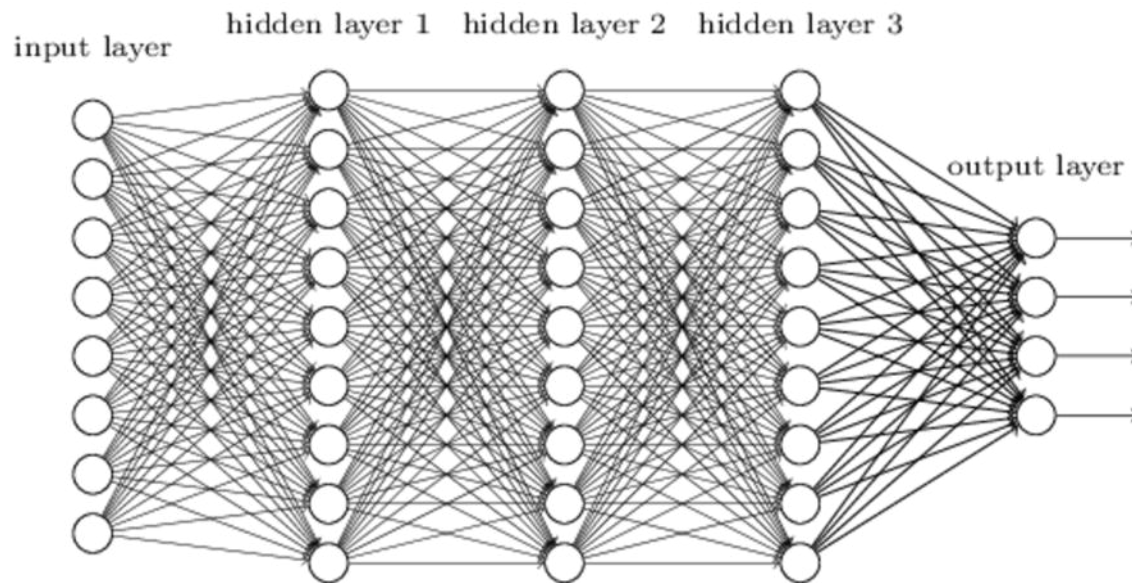
I pesi consentono alla rete di apprendere e catturare le relazioni complesse tra i dati di input e gli output desiderati.

Durante la fase di addestramento, i pesi vengono regolati iterativamente 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.

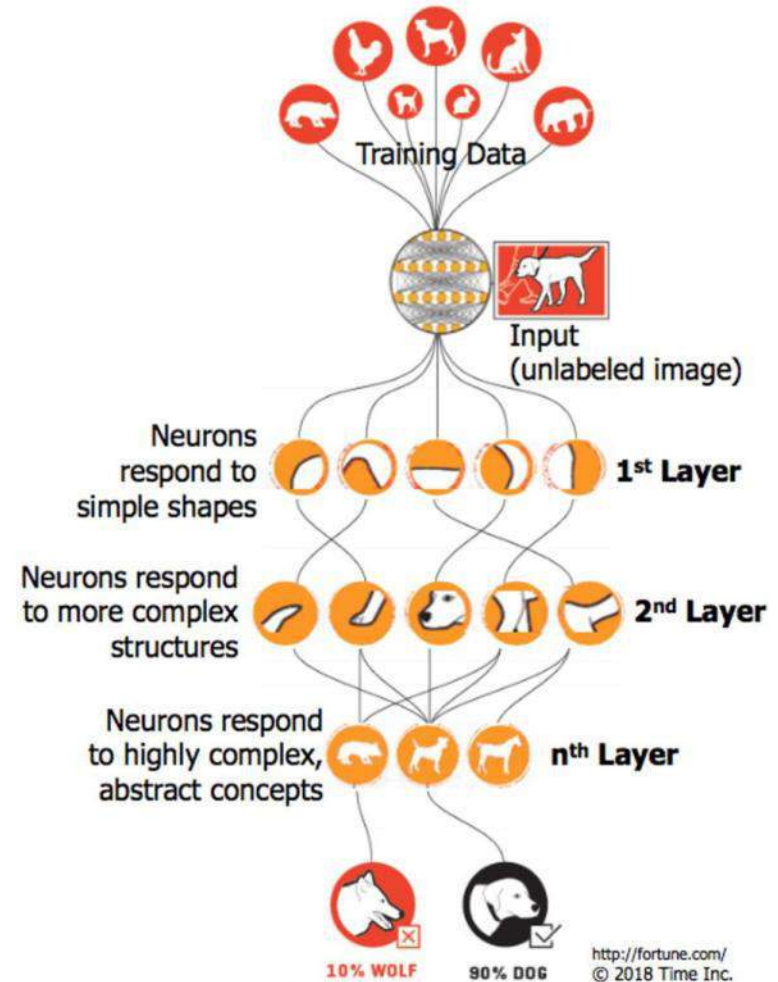
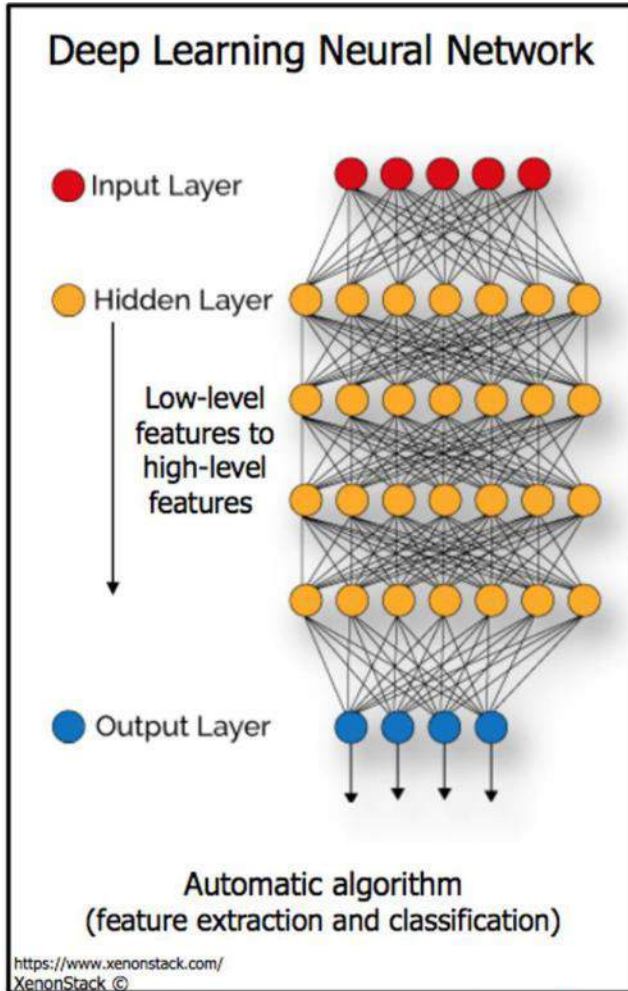


Deep Learning e reti neurali convoluzionali

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.

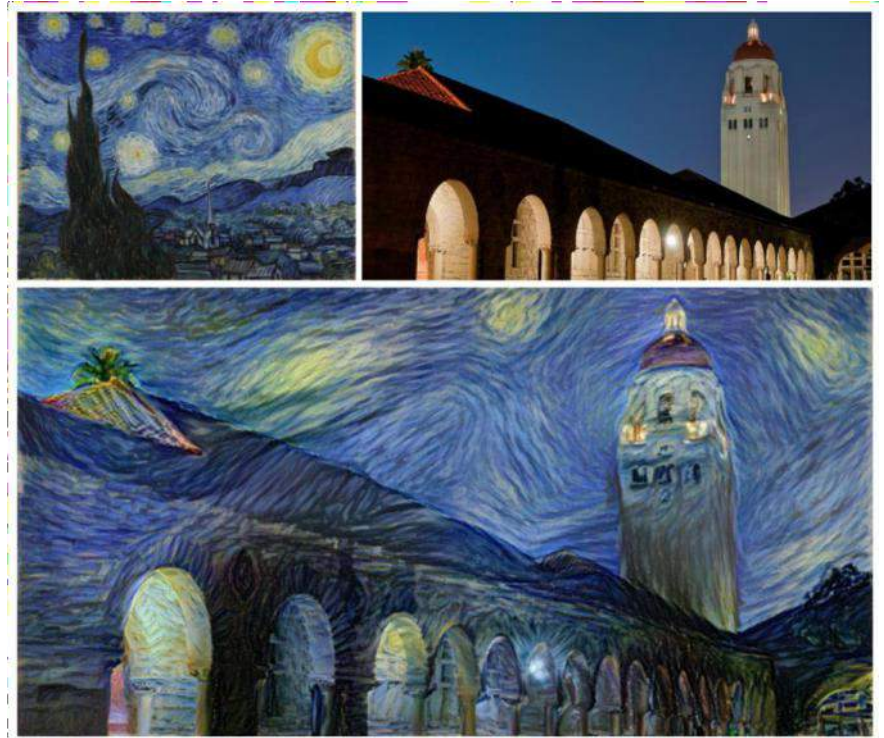


Intuizione di una rete neurale convoluzionale



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>

Deep learning - Arte generativa

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

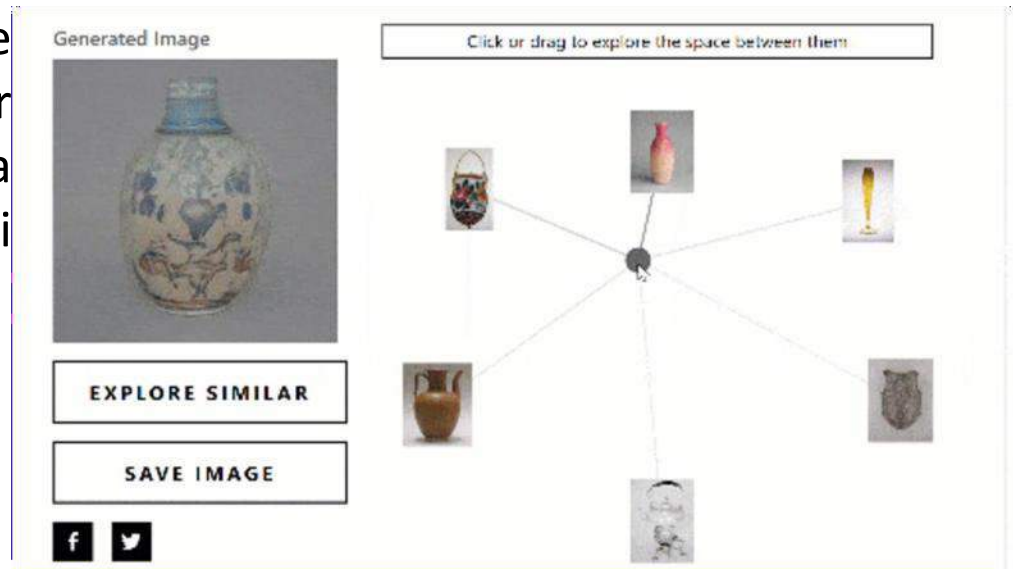
In particolare Gen Studio permette di:

- Estrarre la struttura delle opere d'arte presenti al MET;
- Interpolare tra le caratteristiche delle opere per generare nuovi oggetti.

Fonte:

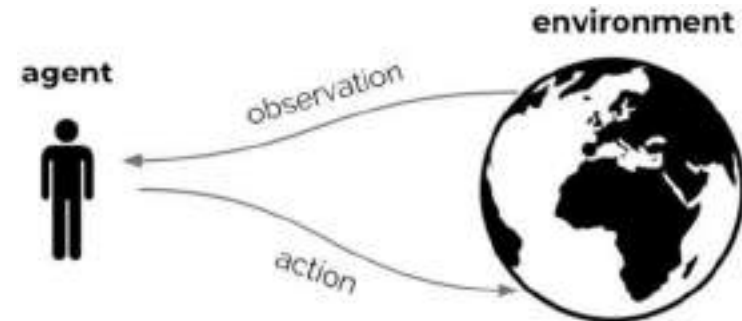
<https://github.com/Microsoft/GenStudio>

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



Reinforcement learning

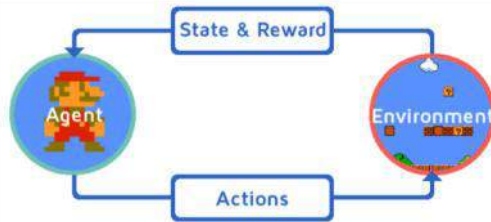
- Noi, e gli altri essere viventi, impariamo dall'interazione con il nostro ambiente
- Questa modalità è diversa rispetto ad altre tipologie di learning
 - Attiva e non passiva
 - Interazione è sequenziale, le future interazioni possono dipendere da quelle precedenti
- Nella nostra interazione con l'ambiente 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



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).

Quando usarlo

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. Questo tipo di tecniche, fondamentalmente andando per tentativi, imparano un comportamento che lo porti con successo all'obiettivo.

Reinforcement learning

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.

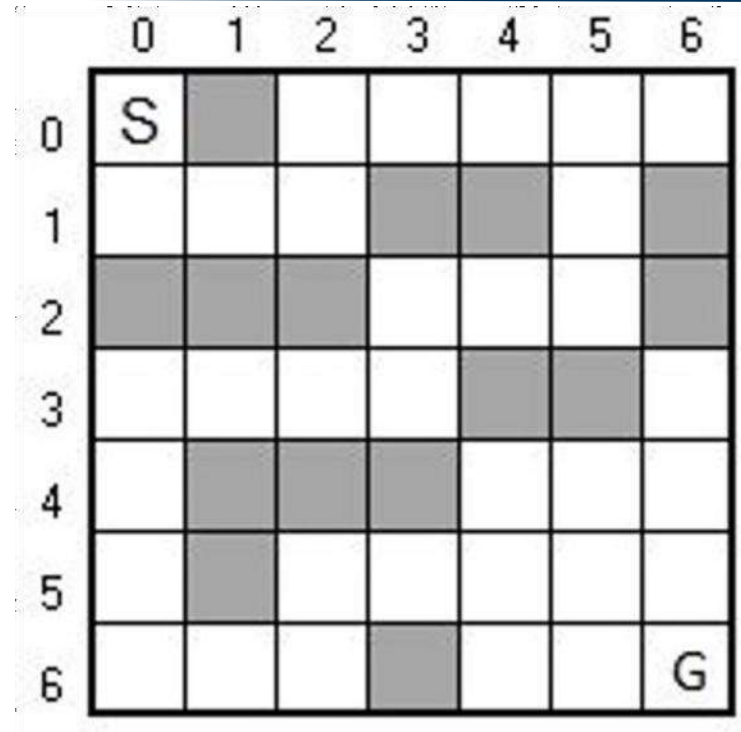


Reinforcement learning - esempio

Esempio: un robot deve uscire da un labirinto

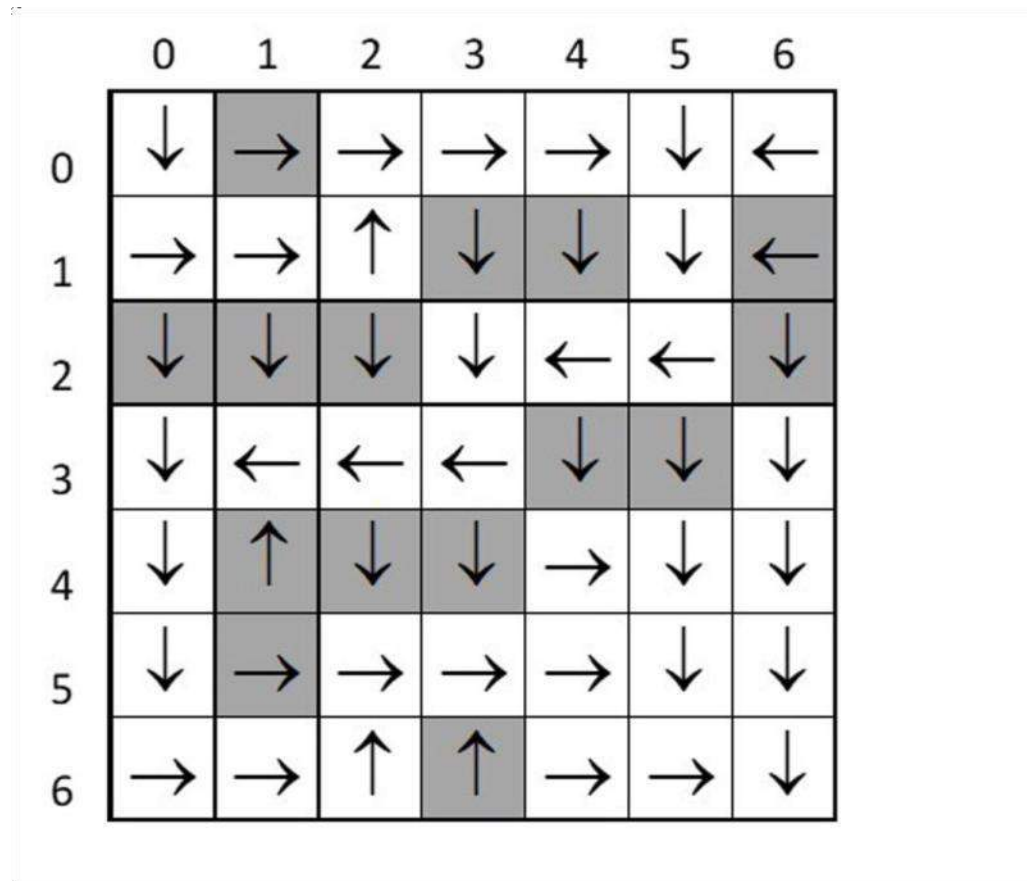
Per incoraggiare il robot a trovare il percorso più breve, viene applicata una piccola penalità di 0,04 unità ogni volta che il robot si sposta in una cella vuota (bianca) e gli ostacoli sono posti intorno al labirinto (contrassegnati in grigio) che provocano una penalità maggiore di 0,75 unità se il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



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.**

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".

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.

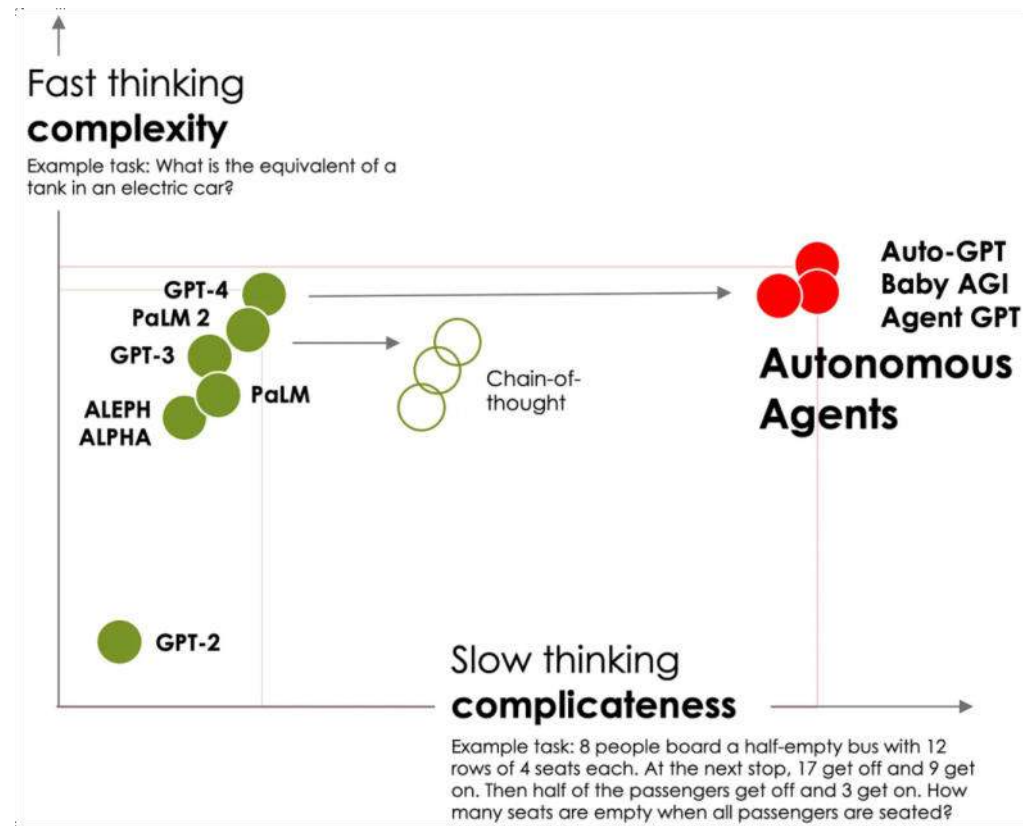
Tuttavia, **l'autonomia di ChatGPT è limitata dalla sua stretta attenzione all'elaborazione e alla comprensione del linguaggio.** Può solo "prendere decisioni" nell'ambito della generazione di testo pertinente in base all'input e **non possiede la capacità di adattarsi o apprendere in modo indipendente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

Autonomous agents (2)

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:

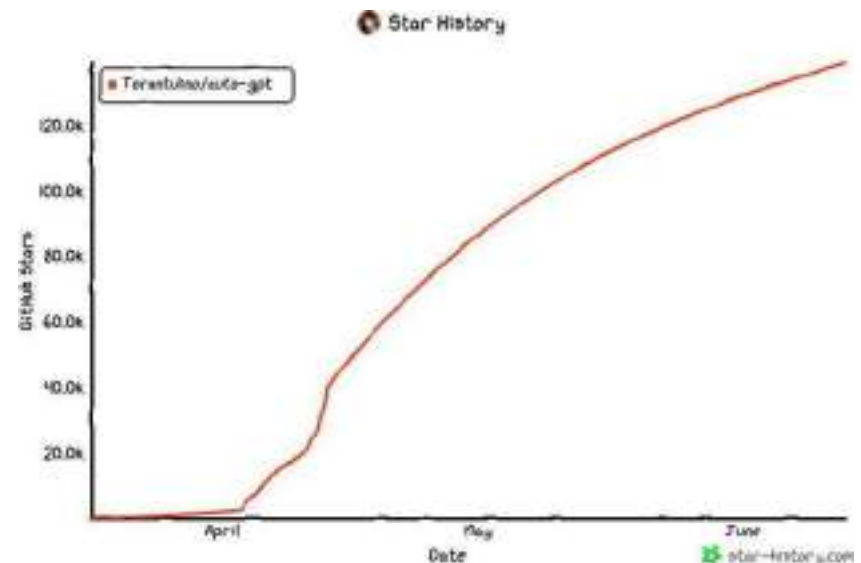
Shen, Y., Song, K., Tan, X., Li, D., Lu, W., & Zhuang, Y. (2023). Hugginggpt: Solving ai tasks with chatgpt and its friends in huggingface. *arXiv preprint arXiv:2303.17580*.

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*

AutoGPT

Auto-GPT è un agente intelligente capace di scomporre un obiettivo assegnatogli in linguaggio naturale e di dedurne 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>

AgentGPT

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 hotel 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)

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.

Normativa

Libro bianco sull'intelligenza artificiale

“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>

Le sfide per l'IA al servizio del cittadino

- **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>

IA e GDPR (Regolamento n. 2016/679)

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.

Riflessioni legate alla IA nel 2020

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 ?

Digital Service Act e AI

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."

La legge Europea - AI ACT

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.

- 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>

AI ACT - misure di trasparenza

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.

Strategia di ENISA per la AI

L'Agenzia 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/>

ECAT - Centro europeo per la trasparenza algoritmica

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.

Segnali di allarme

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

I risultati spaziavano dal riconoscimento di immagini a quello dei volti, 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.**



Segnali di allarme (2)

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>



AI nella guerra in Iran, Anthropic fa causa al Pentagono

Dopo i dissidi sull'utilizzo di Claude nella guerra scoppiata in Iran, la vicenda finisce in tribunale

www.ilsole24ore.com

<https://www.ilsole24ore.com/art/ai-guerra-iran-anthropic-fa-causa-pentagono-AI1JWJrB>

06:59



Un'azienda licenzia tutti i dipendenti a Venezia, sostituiti dall'AI - Notizie - Ansa.it

InvesCloud Italy, 'modello di business non più sostenibile' (ANSA)

www.ansa.it

[https://www.ansa.it/sito/notizie/cronaca/2026/03/11/unazienda-licenzia-tutti-i-dipendenti-a-veneziasostituiti-dallai_204eb62a-eae2-4d61-94e7-40e0ce963804.html?](https://www.ansa.it/sito/notizie/cronaca/2026/03/11/unazienda-licenzia-tutti-i-dipendenti-a-veneziasostituiti-dallai_204eb62a-eae2-4d61-94e7-40e0ce963804.html?fbclid=IwB21leAQeQfxjbGNrBB5B9mV4dG4DYWVtAjExAHNydgMGYXBwX2lkDDM1MDY4NTUzMTcyOAABHpXHaTb5h57WjOwmeI4qIS1b_JQIMJBEaRVLBGoH2WbDWNuDo3AnHLd4cEXV_aem_AM0FaQ2trG1hKqSexoLusa)

[fbclid=IwB21leAQeQfxjbGNrBB5B9mV4dG4DYWVtAjExAHNydgMGYXBwX2lkDDM1MDY4NTUzMTcyOAABHpXHaTb5h57WjOwmeI4qIS1b_JQIMJBEaRVLBGoH2WbDWNuDo3AnHLd4cEXV_aem_AM0FaQ2trG1hKqSexoLusa](https://www.ansa.it/sito/notizie/cronaca/2026/03/11/unazienda-licenzia-tutti-i-dipendenti-a-veneziasostituiti-dallai_204eb62a-eae2-4d61-94e7-40e0ce963804.html?fbclid=IwB21leAQeQfxjbGNrBB5B9mV4dG4DYWVtAjExAHNydgMGYXBwX2lkDDM1MDY4NTUzMTcyOAABHpXHaTb5h57WjOwmeI4qIS1b_JQIMJBEaRVLBGoH2WbDWNuDo3AnHLd4cEXV_aem_AM0FaQ2trG1hKqSexoLusa)

Theoretical capability and observed usage by occupational category

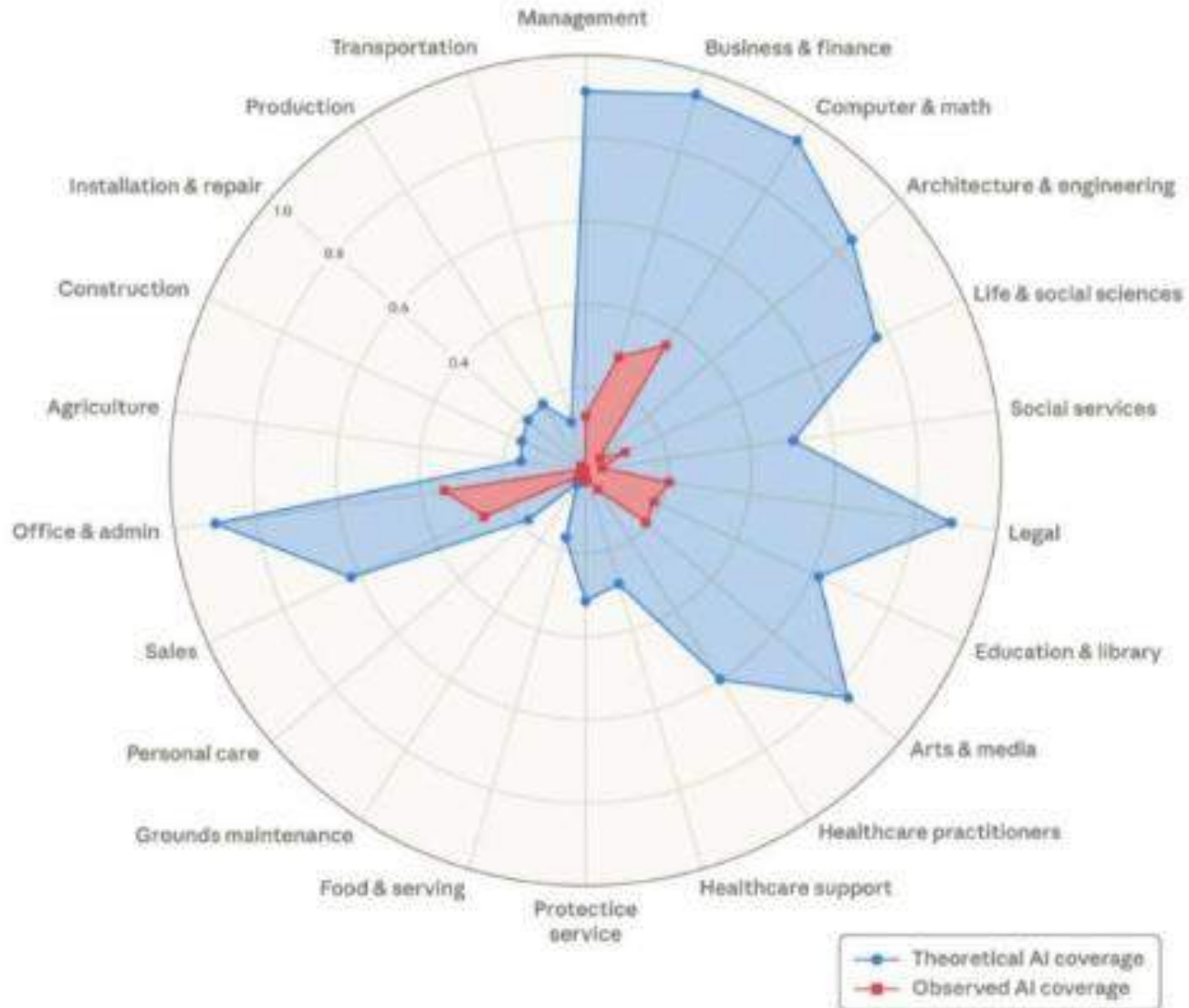


Figure 2: Theoretical capability and observed exposure by occupational category

This figure shows the share of job tasks that LLMs could theoretically perform (blue area) and our own job coverage measure derived from usage data (red area).

Molti nuovi tools online ogni giorno

FUTUREPEDIA Favourites Discover Submit Community Login

Try our new AI search to find the best AI tools!

150K+ 350K+ Sponsor Us

FUTUREPEDIA

THE LARGEST AI TOOLS DIRECTORY, UPDATED DAILY

Tools Added Today 28 News Added Today 4

What do you want to do?

Filters 3D art audio editing avatars code assistant copywriting customer support Sort By Verified

[VIEW ALL CATEGORIES](#)

Artificial Intelligence powered Ad & Social Creatives \$29/mo

The Ultimate Video Co-pilot for Marketers Produce videos 10x faster with the power of generative AI

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

La ricerca non si arresta

Il 26 maggio 2023 FDA americana ha approvato la sperimentazione umana per i prodotti di neuralink che si occupa di interfacce neurali impiantabili



Con l'AI generativa la pubblicità entra all'interno del contenuto

Esiste una regola non scritta nell'economia digitale: ogni servizio che non richiede un pagamento monetario diretto non è privo di costo, ma semplicemente nasconde la forma in cui il costo viene riscosso. Non si tratta di una metafora, la generazione di valore implica sempre consumo di risorse: computazionali, energetiche, umane.

Il meccanismo di base è quello descritto dalla legge di Metcalfe: il valore di una rete cresce proporzionalmente al quadrato del numero dei suoi utenti. Per una piattaforma, questo significa che il modo più razionale di crescere è abbassare al massimo le barriere di ingresso, accettando di non monetizzare nell'immediato.

In questo schema esisteva ancora una distinzione riconoscibile: il contenuto che l'utente sceglie di vedere e l'annuncio pubblicitario che è costretto a subire sono due entità separabili. L'una interrompe l'altra, ma non ne modifica la sostanza. Lo spot televisivo non riscrive la sceneggiatura del film.

Tale logica non è trasferibile all'AI generativa, che rappresenta un cambiamento qualitativo rispetto a tutti i modelli precedenti. Per capire perché, è necessario partire dai costi. Il costo marginale per restituire un risultato su un motore di ricerca tradizionale è tecnicamente trascurabile: recuperare e ordinare pagine già indicizzate richiede risorse computazionali minime. Al contrario, generare un testo tramite un'AI è un'operazione con un consumo energetico che non scala in modo lineare con la crescita degli utenti.

Mantenere questi sistemi richiede una spesa operativa che non può essere coperta dalla sola raccolta di dati comportamentali: quindi, le aziende si trovano quindi a dover scegliere tra far pagare l'utente o l'inserzionista.

La seconda opzione apre un problema di manipolazione semantica che non ha precedenti nella storia della pubblicità digitale. A differenza di quanto avviene oggi, dove la pubblicità è un corpo logicamente estraneo rispetto al contenuto, nell'AI generativa essa non è più separabile dal contenuto, perché diventa parte integrante del processo generativo stesso. La pubblicità sarà incorporata direttamente nella logica con cui il sistema costruisce le proprie risposte, andando ben oltre i cosiddetti "branded content", dove la comunicazione del brand è ben presente ma mantiene una chiara identificabilità.

Il meccanismo tecnico di implementazione è sottile. In un modello finanziato da inserzionisti, i pesi delle connessioni neurali possono essere alterati durante la fase di addestramento o di allineamento per favorire specifici risultati commerciali. La differenza non è percepibile dall'esterno, poiché la risposta mantiene la stessa forma sintattica e lo stesso registro argomentativo.

L'utente non ha strumenti per distinguere se la risposta è frutto di un calcolo probabilistico

calibrato sull'utilità o di una funzione di costo ottimizzata per il profitto dell'inserzionista. Non c'è uno spot che interrompe il ragionamento: lo contamina dall'interno. La provocazione lanciata in uno spot recentemente trasmesso durante il Superbowl rappresenta bene come questo sistema potrebbe evolvere in futuro. Le implicazioni di questo scenario investono direttamente la struttura cognitiva della sfera pubblica. Si sta delineando una stratificazione a due velocità non già nell'accesso all'informazione, ma nella qualità del processo con cui l'informazione viene elaborata e restituita.

Da una parte si collocheranno coloro che possono permettersi di pagare un abbonamento per sistemi che operano con l'obiettivo dichiarato di servire l'utente. Il vantaggio competitivo che ne deriva non si misura solo in velocità di elaborazione, ma in profondità e affidabilità del ragionamento assistito. Dall'altra parte ci sarà la maggioranza degli utenti, che accederà a sistemi apparentemente identici nei formati e nel registro comunicativo, ma strutturalmente diversi nella funzione: non agenti al servizio dell'utente, ma agenti commerciali mascherati da assistenti neutrali. Queste persone avranno accesso a una conoscenza curata e orientata, che produce l'illusione dell'autonomia cognitiva mentre la erode sistematicamente.

Accettare la gratuità in questo contesto significa accettare di non essere i committenti del proprio agente cognitivo, ma il prodotto che esso vende a terzi. La sovranità cognitiva ha un costo, ma è la condizione necessaria per non delegare la propria capacità di giudizio a un sistema il cui padrone non è l'utente che lo interroga. Sotto un profilo politico, la tecnologia non è semplicemente un mezzo rispetto ai fini: diventa un fattore che ridistribuisce asimmetricamente la capacità di formarsi giudizi fondati, con effetti che investono direttamente le condizioni di possibilità del discorso democratico.

— Massimo Chiriatti

Chief Technology & Innovation Officer, Italy, Leaito

© INNOVAZIONE DIGITALE

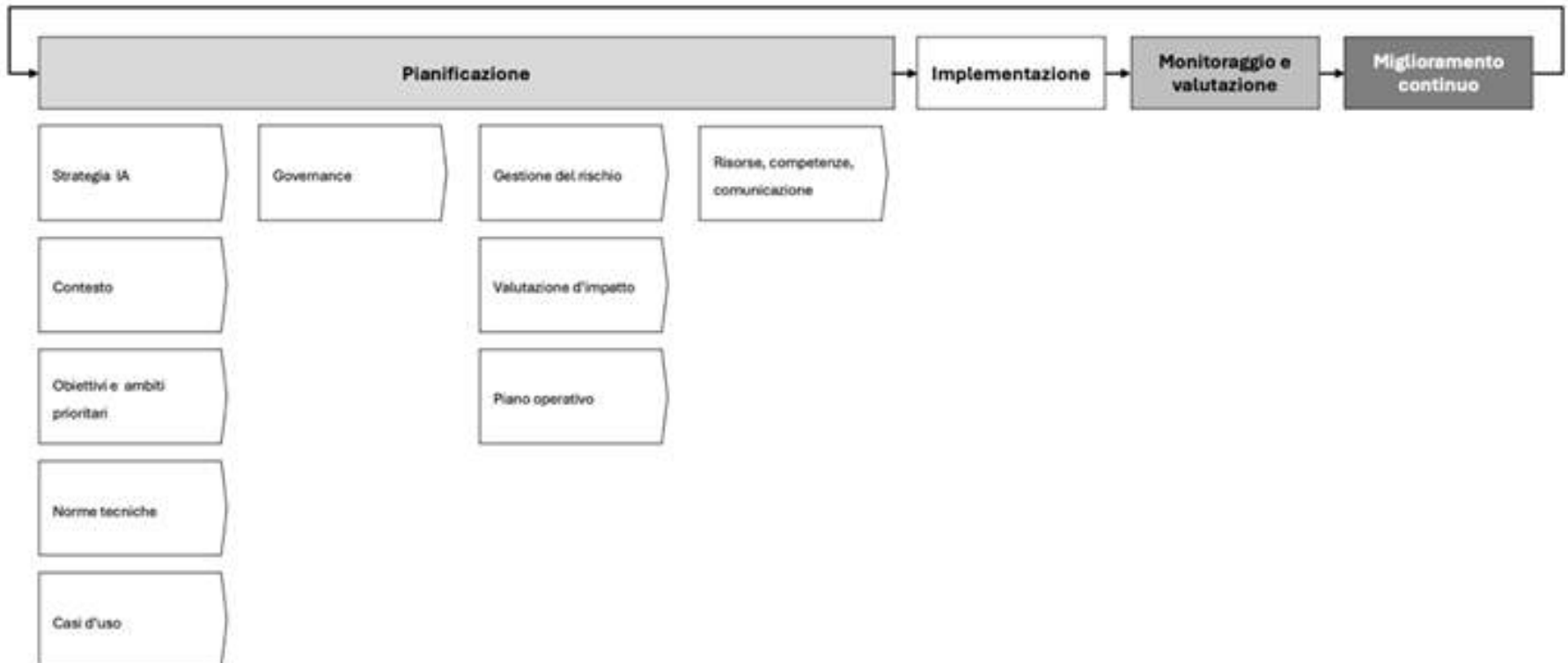
Modello di adozione dell'IA nella PA

Le PA DOVREBBERO adottare un insieme strutturato di processi, politiche, risorse e strumenti per governare, implementare, monitorare e migliorare l'utilizzo dei sistemi di IA durante il loro ciclo di vita.

Modello di adozione dell'IA nella PA

Le PA DOVREBBERO implementare un modello di adozione dell'IA in grado di rispondere prontamente ai cambiamenti del contesto normativo e tecnologico. Le presenti Linee guida propongono un modello di adozione dell'IA basato sul ciclo per il miglioramento continuo Plan-Do-Check-Act (PDCA) e su alcune pratiche gestionali definite dallo standard ISO/IEC 42001:202313. La Figura riporta una rappresentazione sintetica del modello, approfondito nei successivi paragrafi.

Modello di adozione dell'IA nella PA



Modello di adozione dell'IA nella PA

Le PA POSSONO adottare tale modello adattandolo alle proprie caratteristiche, esigenze e responsabilità, con particolare riferimento al ruolo di fornitore o deployer di sistemi di IA. In particolare, le PA POSSONO, per l'adozione di sistemi di IA di rischio limitato, minimo o nullo, accorpare le fasi del modello, adeguandole alla bassa complessità del progetto di adozione.

Strategia per l'IA

La PA DEVE sviluppare una strategia per l'IA coerente con il proprio contesto e allineata alla propria missione in termini di funzione amministrativa, giurisdizione e ambiti operativi specifici.

Le PA POSSONO definire una strategia comune in base alle proprie caratteristiche e tipologia (ad esempio comuni o università).

La strategia DEVE contenere gli obiettivi dell'utilizzo dell'IA e le azioni per conseguirli, promuovendo un approccio condiviso e collaborativo che coinvolga tutte le funzioni della PA.

Strategia per l'IA

Le PA includono nella strategia per l'IA le azioni per:

- **migliorare la qualità dei dati, ivi comprese le basi documentali;**
- **migliorare le competenze del personale interno in materia di IA e di protezione dei dati personali coinvolti;**
- **individuare i casi d'uso che possono apportare maggiore beneficio;**
- **valutare le esperienze e sperimentazioni già effettuate da altre PA;**
- **avviare sperimentazioni, anche in forma associata con altre PA, partendo dai casi d'uso di minore complessità (quick win);**
- **avviare collaborazioni con altre PA per la sperimentazione e successivamente l'adozione, acquisto e sviluppo dei sistemi di IA in forma associata.**

Strategia per l'IA

La strategia per l'IA DEVE essere coerente con la strategia definita nel PIAO, la strategia sui dati e il Piano triennale per l'informatica nella Pubblica Amministrazione. Ciò in modo da delineare una strategia completa e coerente con la politica complessiva sulle tecnologie della PA, che coinvolga anche l'eventuale territorio di riferimento.

La PA DEVE implementare la propria strategia dell'IA affidandone il coordinamento, unitamente alla strategia sui dati, al Responsabile per la transizione al digitale (RTD) e al suo ufficio (UTD), come previsto dal piano triennale per l'informatica nella PA 2024-2026, coinvolgendo altresì il Responsabile della protezione dei dati (RPD/DPO).

Analisi del contesto e delle caratteristiche della PA

Le PA DEVONO analizzare i fattori esterni e interni che influenzano le capacità dell'ente di conseguire i risultati attesi dall'uso dei sistemi di IA, con l'obiettivo di individuare una strategia e un modello di adozione adeguati. **Le PA POSSONO** definire strategia e modello in collaborazione con altre amministrazioni o adottando strategie e modelli definiti da enti sovraordinati o della stessa tipologia.

I fattori esterni includono i requisiti normativi definiti da CAD, AI Act, DGA, GDPR e NIS2, oltre agli eventuali requisiti espressi dagli stakeholder esterni (cittadinanza, imprese, altre PA).

I fattori interni comprendono i requisiti relativi a: struttura organizzativa (incluse le dimensioni dell'ente), stakeholder interni (es. personale dell'ente), ambiti operativi specifici, contesto territoriale e capacità tecnologiche. Queste ultime comprendono la disponibilità di dati, infrastrutture e competenze necessarie per l'implementazione e la gestione efficace dell'IA.

Analisi del contesto e delle caratteristiche della PA

Sulla base delle analisi condotte a livello internazionale da organismi quali EU JRC14 e OECD15 è stato definito uno strumento pratico per determinare il livello di maturità organizzativa e tecnologica delle PA.

Dall'analisi del contesto e delle caratteristiche della PA derivano le azioni della strategia per l'IA con particolare riferimento al:

- **miglioramento della qualità del patrimonio della PA in termini di dati e documenti digitali;**
- **rafforzamento delle competenze;**
- **definizione di collaborazioni con altre PA per la sperimentazione e la gestione associata dell'IA.**

L'analisi del contesto fornisce gli elementi per valutare la fattibilità degli obiettivi e individuare i casi d'uso in cui l'IA può essere impiegata in modo efficace, definendone le modalità di implementazione.

Obiettivi e ambiti prioritari di applicazione

Le PA DEVONO adottare le tecnologie di IA indentificando in via preliminare gli obiettivi e gli ambiti prioritari di applicazione sulla base del proprio contesto, pur potendo le PA condurre sperimentazioni di uso dell'IA in ambiti non prioritari e meno soggetti a rischi.

Le PA DEVONO verificare le proprie esigenze in modo dettagliato e documentato, con l'obiettivo di identificare i casi d'uso in cui l'IA offre il massimo beneficio, in termini di miglioramento dell'efficienza operativa e dell'erogazione dei servizi.

La verifica delle esigenze specifiche contribuisce all'eventuale studio di fattibilità e alla valutazione comparativa necessaria per la scelta delle soluzioni di IA, come previsto dall'art. 68 del CAD.

Gli ambiti prioritari per l'utilizzo dell'IA da parte della PA deriva da un lavoro di ricognizione e analisi condotto su progetti e sperimentazioni in corso a livello nazionale, europeo e internazionale.

Obiettivi e ambiti prioritari di applicazione

Gli ambiti prioritari individuati in tale ricognizione sono:

- 1. Miglioramento dell'efficienza operativa: le PA POSSONO utilizzare l'IA per aumentare la propria capacità di analisi e gestione dei dati e di automatizzazione dei processi ripetitivi, al fine di semplificare i processi interni, ridurre i tempi operativi e migliorare l'efficienza complessiva. In particolare, le aree in cui si individuano i maggiori benefici sono:**

Obiettivi e ambiti prioritari di applicazione

Supporto alle decisioni: le PA impiegano l'IA per sviluppare modelli predittivi che consentano di adottare decisioni consapevoli e basate sui dati reali, aumentando l'affidabilità e la tempestività delle decisioni.

Ottimizzazione dell'allocazione delle risorse: le PA utilizzano l'IA per distribuire le risorse in modo più efficiente, individuando le priorità e focalizzandosi sulle aree di maggiore necessità, ottimizzando così l'uso delle risorse pubbliche.

Miglioramento della gestione documentale: le PA utilizzano l'IA per automatizzare la classificazione, l'archiviazione e il recupero dei documenti, facilitando la ricerca e riducendo i tempi di gestione.

Miglioramento del supporto giuridico: l'IA supporta le PA nell'analisi normativa e giurisprudenziale, consentendo di elaborare pareri legali più accurati e tempestivi e di monitorare aggiornamenti legislativi rilevanti.

Miglioramento delle procedure di acquisto: le PA adottano l'IA per ottimizzare le procedure di procurement, migliorando sia l'efficienza che la trasparenza del processo di acquisto.

Obiettivi e ambiti prioritari di applicazione

2. Miglioramento dei servizi ai cittadini e alle imprese: le PA POSSONO utilizzare l'IA per aumentare la propria capacità di analisi e gestione dei dati al fine di personalizzare i servizi digitali in base alle specifiche esigenze degli utenti, anche in logica proattiva. In particolare, le aree in cui si individuano i maggiori benefici sono:

Obiettivi e ambiti prioritari di applicazione

- a. **Personalizzazione:** le PA impiegano l'IA per adattare i servizi pubblici alle esigenze specifiche di cittadini e imprese, migliorando l'interazione digitale e l'efficienza nella risposta.
- b. **Proattività:** l'IA consente alle PA di anticipare le esigenze degli utenti, fornendo servizi o informazioni pertinenti prima che siano richiesti, semplificando così l'accesso e riducendo i tempi di attesa.
- c. **Trasparenza:** le PA utilizzano l'IA per migliorare la trasparenza, fornire a cittadini e imprese informazioni chiare e immediatamente fruibili sui propri adempimenti e sullo stato di avanzamento dei procedimenti amministrativi avviati presso la PA stessa.
- d. **Accessibilità:** le PA adottano soluzioni di IA per rendere i servizi pubblici accessibili e conformi all'art. 53 del CAD, garantendo l'usabilità delle piattaforme digitali anche a persone con disabilità o con limitate competenze digitali. In particolare, l'IA deve essere utilizzata come strumento di assistenza per la creazione e la gestione di contenuti nativamente accessibili.
- e. **Inclusione:** le PA adottano l'IA per analizzare le esigenze dei cittadini al fine di promuovere servizi destinati alle fasce deboli della popolazione, promuovendo l'inclusione sociale.

Obiettivi e ambiti prioritari di applicazione

3. Sicurezza e protezione dei dati:

le PA POSSONO utilizzare l'IA per migliorare la sicurezza dei dati e delle infrastrutture, identificando potenziali minacce e garantendo una protezione avanzata.

Casi d'uso

Considerato quanto detto ai paragrafi precedenti, il modello proposto per l'adozione dell'IA prevede che le PA individuino i casi d'uso in cui l'IA offre il massimo beneficio in termini di miglioramento dell'efficienza operativa e dell'erogazione dei servizi, sulla base della propria strategia per l'IA , dell'analisi del contesto e delle caratteristiche dell'organizzazione , tenuto conto degli obiettivi e gli ambiti prioritari di applicazione dell'IA .

Le PA POSSONO comunque identificare, operando una semplificazione del modello, i casi d'uso specifici per sperimentazioni, progetti pilota e iniziative circoscritte a basso rischio.

Una volta identificati i casi d'uso, le PA DEVONO raccogliere, documentare e aggiornare le informazioni fondamentali sugli stessi e sulle soluzioni di IA che li implementano, durante l'intero ciclo di vita di queste ultime, dall'ideazione alla prova di concetto sperimentale (POC), al rilascio in ambiente operativo, fino alla dismissione.

I casi d'uso POSSONO essere implementati in forma associata da più PA, anche attraverso gli spazi di sperimentazione e sviluppo previsti dal piano triennale per l'informatica nella PA.

Nel caso di sviluppo in forma associata, la gestione della documentazione è affidata alla PA capofila o a una delegata.

Le PA DEVONO trasmettere periodicamente all'AgID un estratto della documentazione descrittiva dei casi d'uso secondo le modalità definite nel Piano triennale per l'informatica nella pubblica amministrazione.

Requisiti

Nelle fasi di sviluppo e gestione dei sistemi di IA, le PA DEVONO identificare i requisiti a partire dalle esigenze normative, tecniche, etiche e operative che devono essere soddisfatte, tenendo presente le aspettative delle parti interessate.

Le PA DEVONO tenere in considerazione, oltre ai Principi indicati, le seguenti categorie di requisiti:

- **conformità all'AI Act**
- **gestione e qualità dei dati**
- **protezione dei dati personali**
- **sicurezza cibernetica .**

Nell'identificazione dei requisiti, le PA DEVONO tenere in considerazione anche il livello di rischio del sistema di IA da implementare: nel caso di sistemi ad alto rischio, si applicano i requisiti obbligatori previsti dell'AI Act.

Indicatori di prestazione

Le PA DEVONO definire, per ciascun caso d'uso, indicatori di prestazione (*Key Performance Indicator, KPI*) che misurino l'efficacia del sistema di IA nel raggiungimento degli obiettivi prefissati.

I KPI POSSONO essere differenti a seconda del livello di maturità del sistema di IA.

I KPI DEVONO essere monitorati nel corso dell'evoluzione del caso d'uso il ciclo di vita consentendo di determinare l'evoluzione delle prestazioni del sistema di IA nelle sue diverse fasi di maturità.

I KPI POSSONO misurare sia prestazioni tecniche sia l'impatto sui processi aziendali e il valore aggiunto per la PA.

Indicatori di prestazione

I KPI possono riguardare:

- **prestazioni del modello;**
- **qualità dei dati;**
- **robustezza e affidabilità;**
- **efficienza computazionale;**
- **usabilità e accessibilità;**
- **impatto etico e conformità normativa;**
- **costo e sostenibilità.**

Gli organi direttivi della PA approvano la strategia per l'IA dell'ente.

Secondo quanto previsto dal Piano triennale per l'informatica nella pubblica amministrazione, la responsabilità della gestione dell'IA nelle PA deve essere affidata al RTD e al suo ufficio (UTD).

Gli organi direttivi DEVONO garantire questa assegnazione e la devono comunicare all'interno della PA.

Nello specifico, al RTD è assegnata la responsabilità e l'autorità per:

- **assicurare che i sistemi di IA siano conformi alle indicazioni delle presenti Linee guida;**
- **referire agli organi direttivi sulle prestazioni e sulle attività di controllo, monitoraggio e evoluzione dei sistemi di IA.**

La PA DOVREBBE definire procedure di governance per lo sviluppo e l'utilizzo dei sistemi di IA. Le eventuali procedure DEVONO essere coerenti con la strategia per IA dell'ente e con la strategia definita nel PIAO.

Fermi restando gli obblighi di conformità all'AI Act, le PA DEVONO adottare un codice etico per l'IA.

Tale codice DEVE divenire uno strumento di governance vincolante, allineato con il quadro normativo vigente, integrato nei processi decisionali e operativi della PA, finalizzato a un uso responsabile, equo e trasparente dell'IA

Piano Operativo

Le PA DEVONO definire obiettivi operativi chiari per l'adozione e l'uso dell'IA assicurandone la coerenza con la propria strategia per l'IA e il rispetto dei requisiti normativi.

Gli obiettivi DEVONO essere misurabili, monitorati e aggiornati periodicamente.

Per il raggiungimento degli obiettivi, la PA DEVE adottare un approccio sistematico e documentato di project management definendo un piano operativo, nel quale DEVE:

- **definire le attività operative che devono essere effettuate;**
- **identificare le risorse richieste;**
- **definire i tempi entro i quali le attività devono essere completate;**
- **assegnare responsabilità specifiche;**
- **stabilire criteri di valutazione e monitoraggio dei risultati.**

Il piano operativo DEVE includere tutte le attività necessarie per il conseguimento degli obiettivi riguardanti la gestione , il procurement , lo sviluppo dei sistemi di IA adottati o adottandi dalla PA.

Risorse, competenze, comunicazione

Le PA DEVONO determinare e allocare le risorse finanziarie, tecnologiche e umane adeguate a sviluppare, mantenere e migliorare con continuità i propri sistemi di IA.

Le risorse possono includere:

- **dati utilizzati nelle varie fasi del ciclo di vita del sistema di IA;**
- **algoritmi e modelli di IA;**
- **infrastrutture IT (es.: cloud computing, edge computing, risorse di elaborazione);**
- **risorse umane con le competenze necessarie a gestire il ciclo di vita del sistema di IA.**

Le PA DEVONO determinare le competenze necessarie per il personale che svolge attività relative ai sistemi di IA. Le PA DEVONO garantire che tali persone acquisiscano le competenze necessarie tramite formazione, istruzione o esperienza e che conservino documentazione a comprova delle competenze acquisite.

Le PA DEVONO adottare le misure necessarie per acquisire le competenze e per valutarne l'efficacia .

Le PA DEVONO promuovere un uso responsabile ed efficace dell'IA, assicurando che il personale sia pienamente consapevole dei principi, degli obiettivi e della strategia per l'IA, nonché del codice etico e di comportamento.

Le PA DEVONO definire il piano di comunicazione interno ed esterno relativo ai sistemi di IA adottati.

Fermi restando gli obblighi di conformità all'AI Act, le PA DEVONO documentare e conservare le informazioni necessarie per assicurare la conformità normativa, trasparenza e tracciabilità dei sistemi di IA. È inclusa la documentazione operativa (manuali, procedure, valutazioni di impatto e report di monitoraggio) necessaria a garantire l'efficacia e il miglioramento continuo del sistema.

Le PA applicano a tale documentazione le Linee guida AgID sulla formazione, gestione e conservazione dei documenti informatici.

Conclusione

“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

Antonio Memoli

antonio.memoli.dpo@gmail.com

