Introduzione all'intelligenza artificiale

13 settembre 2023

Andrea Bernardini





Argomenti del giorno

- Contesto
- Introduzione
- Strumenti di Al nel 2023
- Scenari di uso industriali
- Cenni storici e terminologia
- Struttura di applicazioni di Al
- Modelli di Intelligenza Artificiale
- Prossime evoluzioni dell'Al
- Normativa e attività della UE



Fonte/Crediti: Immagine creata con Midjourney.com



Argomenti del giorno

- Contesto
- Introduzione
- Strumenti di Al nel 2023
- Scenari di uso industriali
- Cenni storici e terminologia
- Struttura di applicazioni di Al
- Modelli di Intelligenza Artificiale
- Prossime evoluzioni dell'Al
- Normativa e attività della UE

Ma prima di cominciare qualche domanda



Fonte/Crediti: Immagine creata con Midjourney.com



CONTESTO



Al sta cambiando il settore delle telecomunicazioni

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





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.



OpenAI chief seeks to calm fears on job losses

Fonte/Crediti:

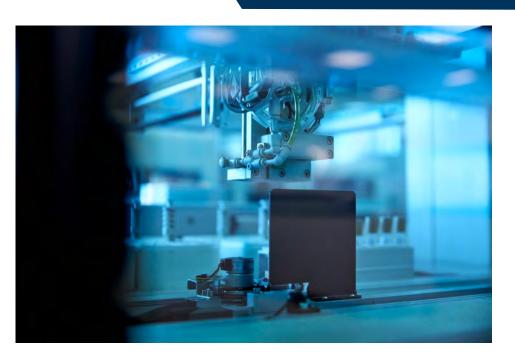
BBC, https://www.bbc.com/news/business-65631168, 18 maggio 2023

Key4Biz, https://www.key4biz.it/bt-tagliera-10-mila-posti-di-lavoro-per-far-posto-allia-entro-il-2030/446818/, 19 maggio 2023

Indiantimes, https://telecom.economictimes.indiatimes.com/news/internet/openai-chief-seeks-to-calm-fears-on-job-losses/100540898, 27 maggio 2023



Anche il settore manifatturiero è in trasformazione



"L'Europa sta guidando il percorso, con più della metà dei suoi principali produttori che implementano almeno un caso d'uso di intelligenza artificiale nelle operazioni di produzione (in Europa, è la Germania a guidare il gruppo, con il 69% dei suoi produttori che implementano l'IA). L'Europa è seguita dal Giappone (30% di implementazione) e dagli Stati Uniti (28%)". Tre casi d'uso si distinguono per la loro idoneità nel dare il via al percorso di intelligenza artificiale di un produttore:

- Manutenzione intelligente
- Controllo della qualità del prodotto
- Pianificazione della domanda

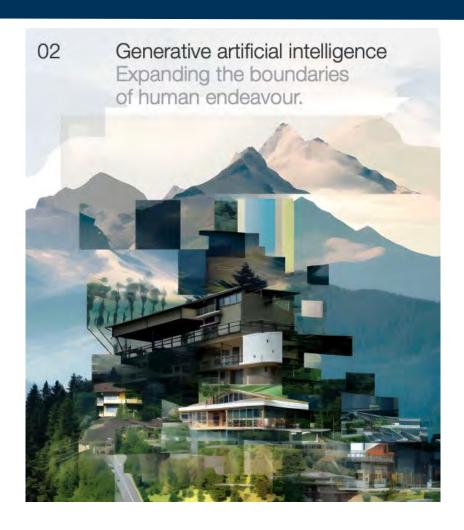
Fonte/Crediti:

Capgemini, https://www.capgemini.com/insights/research-library/scaling-ai-in-manufacturing-operations/ 2019 **Unspalsh**, https://unsplash.com/photos/9xHsWmh3m_4



Generative AI tra le tecnologie emergenti





Fonte/Crediti: World Economic Forum, 2023, https://www3.weforum.org/docs/WEF_Top_10_Emerging_Technologies_of_2023.pdf



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

	The Washington Post Democracy Dies in Darkness				
TECH	Artificial Intelligence	Help Desk	Internet Culture	Space	Tech Policy

INNOVATIONS

ChatGPT took their jobs. Now they walk dogs and fix air conditioners.

Technology used to automate dirty and repetitive jobs. Now, artificial intelligence chatbots are coming after high-paid ones.

By Pranshu Verma and Gerrit De Vynck June 2, 2023 at 6:00 a.m. EDT

Fonte/Crediti:

Repubblica, https://www.repubblica.it/esteri/2023/06/10/news/chatgpt_celebra_messa_germania-403947969/ **Washingtonpost**, https://www.washingtonpost.com/business/2023/09/04/ai-is-chatgpt-coming-for-entry-level-jobs/ https://www.washingtonpost.com/technology/2023/06/02/ai-taking-jobs/



Industrial Digitalization - Al-Enabled Digital Twins

- Digital Product Twins si riferiscono a modelli CAD 3D o rappresentazioni virtuali di beni o prodotti fisici. Le aziende spesso si concentrano solo su questo primo passaggio anziché sulle fasi successive di una soluzione completa di gemello digitale aziendale end-to-end.
- Digital Process Twins determinano il metodo di produzione più efficace per un prodotto. Le aziende simulano processi di produzione utilizzando il gemello digitale per identificare il processo di produzione più adatto in termini di costo, sostenibilità e fattibilità produttiva.
- Digital Factory Twins rappresentano fabbriche virtuali. I dettagli specifici della fabbrica includono costi regionali come la manodopera, le spese per l'elettricità, i costi dei materiali e i tassi di overhead. Possono anche includere le capacità di produzione specifiche per ogni fabbrica.





Fonte/Crediti:

Siemens, https://www.siemens.com/global/en/company/insights/siemens-and-nvidia-partner-to-build-the-industrial-metaverse.html Nvidia, https://resources.nvidia.com/en-us-omniverse-industrial-digital-twins/omniverse-enterprise-5-steps

Apriori, https://www.apriori.com/blog/actionable-digital-transformation-strategies-for-manufacturing/



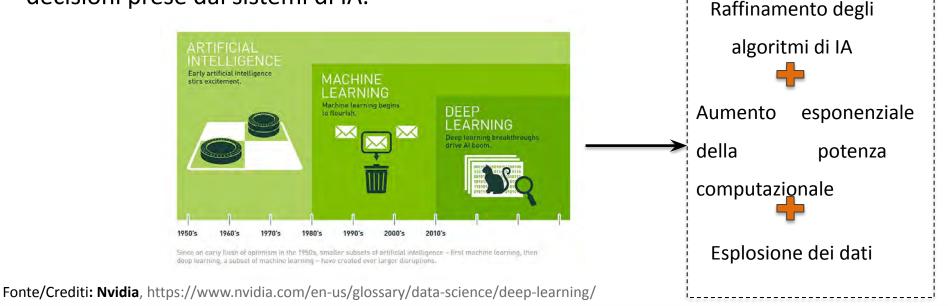
INTRODUZIONE



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.



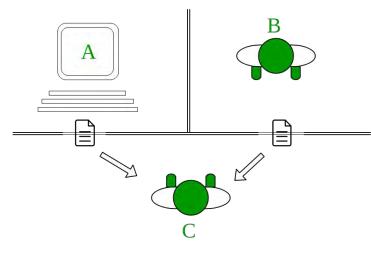


Introduzione - Definizione semplice di IA

"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'intervistatore, 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.

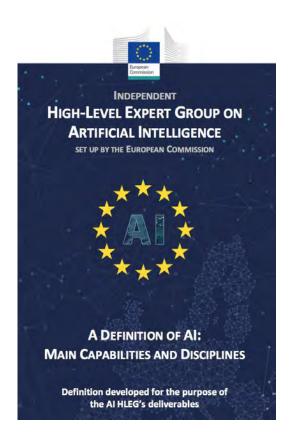
Componenti principali della IA secondo Turing: (linguaggio, ragionamento, conoscenza, apprendimento e comprensione).



Fonte/Crediti: Geekforgeeks, https://www.geeksforgeeks.org/turing-test-artificial-intelligence/



Introduzione - 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/Crediti: European commission

https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines



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

Fonte/Crediti: https://atc.mise.gov.it/images/documenti/Corsi_di_aggiornamento/intelligenza_artificiale.pdf



Settori di utilizzo della Al (debole)

Speech recognition Fraud detection

Autonomous planning and Recommendation systems

scheduling Web search engines

Financial forecasting Autonomous cars

Spam fighting Energy optimization

Logistics planning Question answering systems

Robotics Social network analysis

Machine translation Medical diagnosis

Information extraction Route finding

Automatic assembly Protein design

Sentiment analysis Document summarization

Computer animation Transportation/scheduling



I servizi di Al di cui si parlava nel 2020

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 prodotti e gestione della rete di distribuzione...

Linkedin

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

E tantissimi altri servizi

Ride-sharing apps, smart assistant, etc..



I servizi di Al di cui si parla ad inizio 2023

Google

riconoscimento volti e parlato, classificazione email,

nelle mappe l'odelli di linguaggio di grandi

Facebook

clastification di contenuti offensivi, riconoscimento la Al generativa

contenuti (ChatGPT, AlphaCode Bard, Bing AI, Claude, GitHub Copilot..)

Amazon

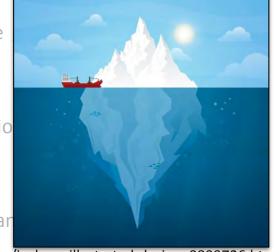
Raccomandazioni di prodotti e distribuzione...

Linkedin

Personalizzazione, comprensio sistemi di raccomandazioni...

E tantissimi altri servizi

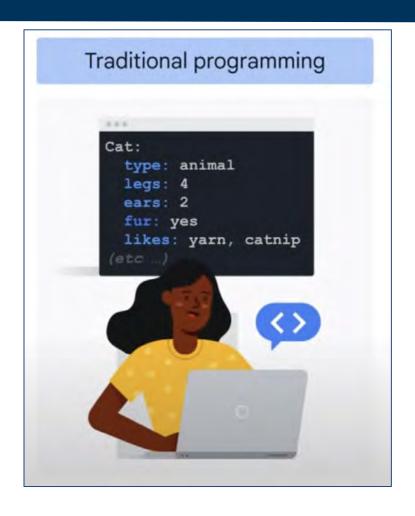
Ride-sharing apps, smart assistar



Fonte/Crediti: https://www.freepik.com/free-vector/iceberg-illustrated-design 9899726.htm

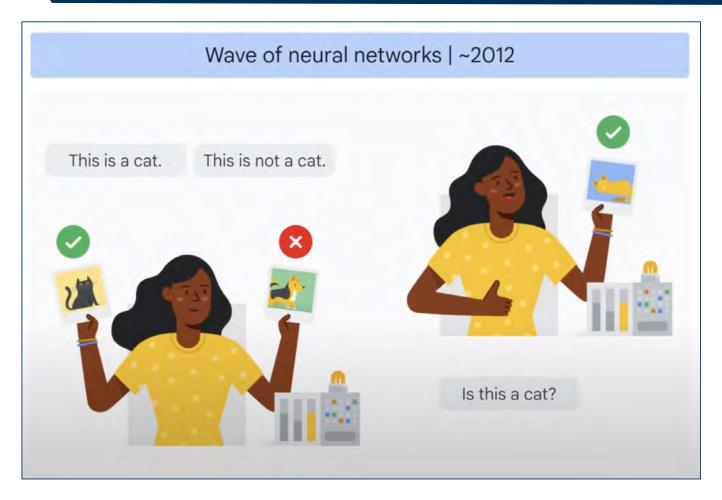


Programmazione tradizionale



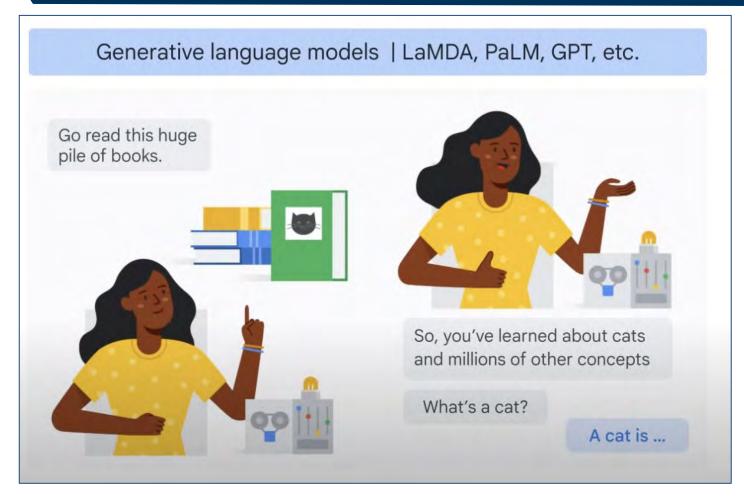


Il ritorno delle reti neurali





I modelli di linguaggio generativi





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



Classify

Discriminative model (classify as a dog or a cat)



Generative technique



Generate

Generative model (generate dog image)





Strumenti di Al nel 2023



ChatGPT

ChatGPT è un chatbot basato su intelligenza artificiale e apprendimento automatico sviluppato da OpenAl 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



Fonte/Crediti: Wikipedia, https://en.wikipedia.org/wiki/ChatGPT



ChatGPT

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

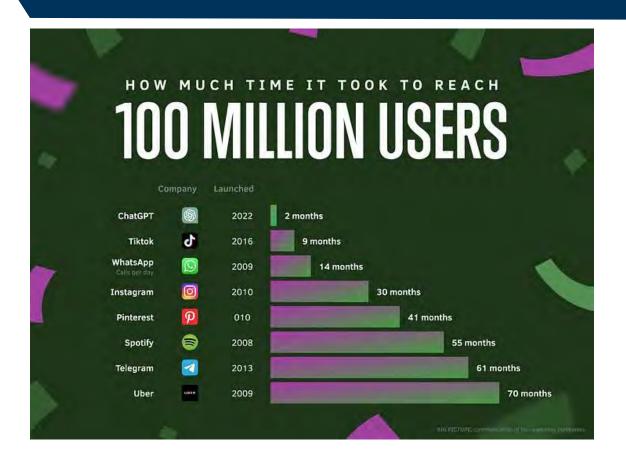
ChatGPT è stato messo a punto con tecniche di machine learning non supervisionato e ottimizzato con tecniche di apprendimento supervisionato e per rinforzo.



Fonte/Crediti: Wikipedia, https://en.wikipedia.org/wiki/ChatGPT



Trend di utilizzo



Fonte/Crediti: **Washington Post**, ChatGPT loses users for first time, shaking faith in AI revolution, Luglio 2023 https://www.washingtonpost.com/technology/2023/07/07/chatgpt-users-decline-future-ai-openai/ **II Sole 24 ore**, Google sbarca con Bard in Italia. Ecco le differenze con ChatGpt, Luglio 2023 https://www.ilsole24ore.com/art/bard-ia-generativa-google-arriva-italia-ecco-differenze-chatgpt-AFuKFUC



ChatGPT

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 ingegneristicamente interessante ma niente di più

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

Noam Chomsky: The False Promise of ChatGPT,

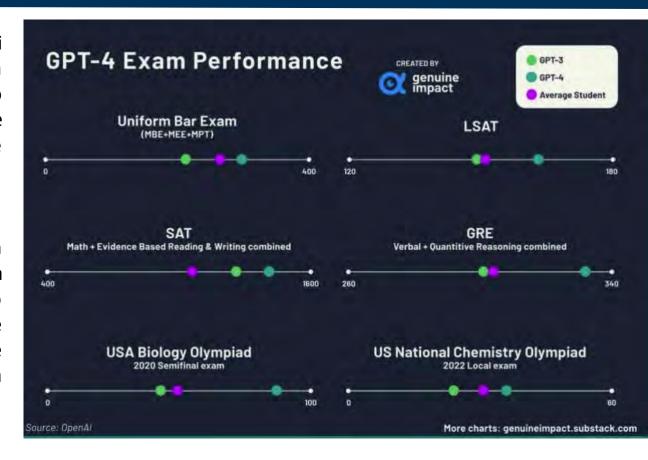
https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html



Capacità di superare test ed esami

Istituzioni come Sciences Po di Parigi e l'Università RV dia Bangalore in India, hanno vietato l'uso di ChatGPT, preoccupate che potesse compromettere l'apprendimento e favorire la frode.

Altre università come la Wharton School of Business e l'Ithaca College a New York l'hanno invece permesso, sostenendo che gli studenti dovrebbero essere competenti nell'uso di questa tecnologia.



Fonte/Crediti:

Washington Post, https://www.washingtonpost.com/technology/2023/08/13/ai-chatgpt-chatbots-college-cheating/

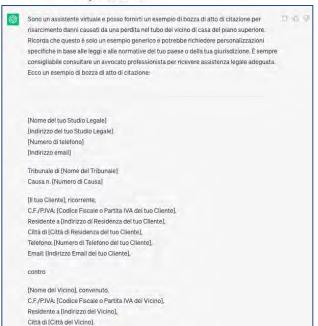
Reddit: https://www.reddit.com/r/singularity/comments/12mjo2f/chatgpt4 exam performances/

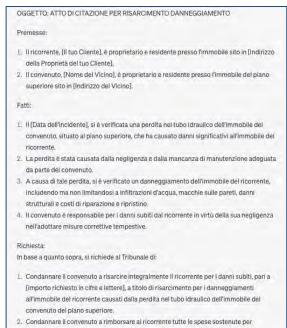


Come usarlo ? (1) - Scrivere documenti



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.





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

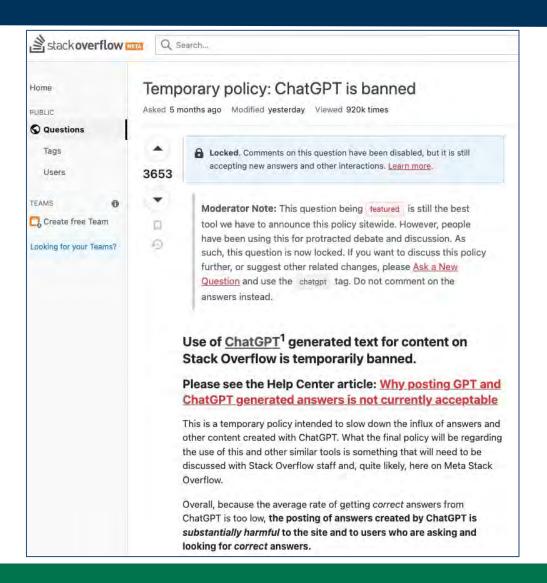
Fonte/Crediti: **Altalex**, https://www.altalex.com/documents/news/2023/01/16/la-chatgpt-che-scriveva-come-un-avvocato Screenshot di interrogazione di ChatGPT



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/Crediti: **Stackoverflow**, https://meta.stackoverflow.com/questions/421831/t emporary-policy-chatgpt-is-banned

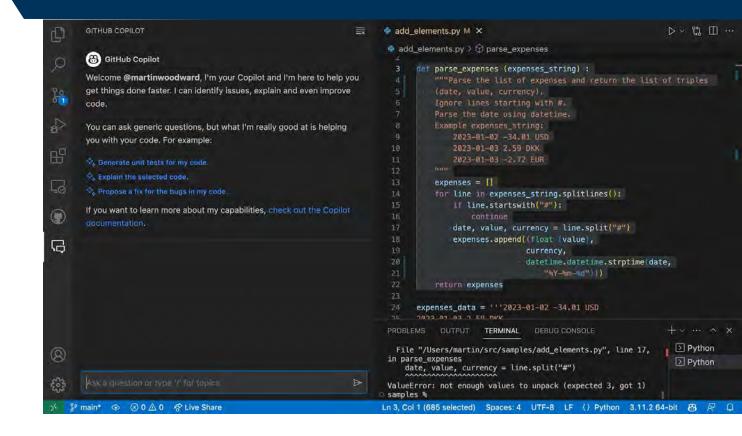




Come usarlo ? (3) - Supporto allo sviluppo

Alcuni utilizzi

- Spiegare codice
- Trovare bug
- Scrivere test
- Scrivere codice



Fonte/Crediti: Github Copilot, https://github.com/features/copilot



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

```
import pandas as pd
from pandasai import PandasAI

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

# Instantiate a LLM
from pandasai.llm.openai import OpenAI
llm = OpenAI(api_token="YOUR_API_TOKEN")
pandas_ai = PandasAI(llm)
pandas_ai = 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/Crediti: Screenshot da https://github.com/gventuri/pandas-ai



Come usarlo ? (5) - Prompt



Fonte/Crediti: Screenshot di interrogazione di ChatGPT

Cyber Security Specialist

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

Programming & Technology

Use prompt →

English Translator and Improver

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

Education & Learning

Use prompt →

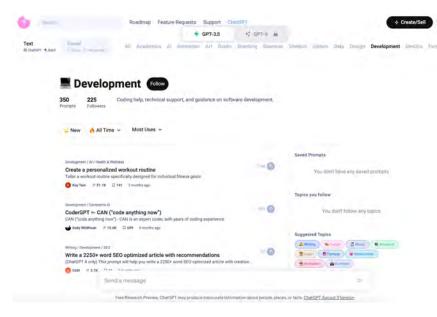
StackOverflow Post

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

Programming & Technology



Come usarlo ? (5) - Prompt



Write a 2250+ word SEO optimized article with recommendations

(ChatGPT 4 only) This prompt will help you write a 2250+ word SEO optimized article with creative title recommendation and SEO ready description, outline and FAQs from your Keywords or Title Text. A creative title and SEO ready description will be written as well. Super important: Don't forget to set the language, tone and style. Enjoy!



As an Expert level SEO Specialist, Content Writer, and Language Expert, you are required to produce an article on the topic "Mkeyword or title" using active voice.

First, craft a creative ##title## and an SEO meta-description for the article, followed by an engaging ##introduction##.

Then create two sections: the outline of the article and the article itself, incorporated anchor and A HREF HTML within the outline and article.

Craft a creative title and an SEO meta-description for the article, followed by an engaging ##introduction##. Bold the heading of the second section using Markdown language.

Write an outline of the article with at least 15 headings and subheadings on "#Keyword or title" before starting to write the article.

Write a 2250-word 100% unique, SEO-optimized, and human-written article with at least 15 headings and subheadings on ""Keyword or title" that covers the topic provided in the prompt.

Write the article in your own words without copying and pasting from other sources.

Consider perplexity and burstiness when creating content, ensuring high levels of both without losing specificity or context.

Use fully detailed paragraphs that engage the reader.

Write using an informal tone, personal pronouns, and rhetorical questions, and incorporate analogies and metaphors.

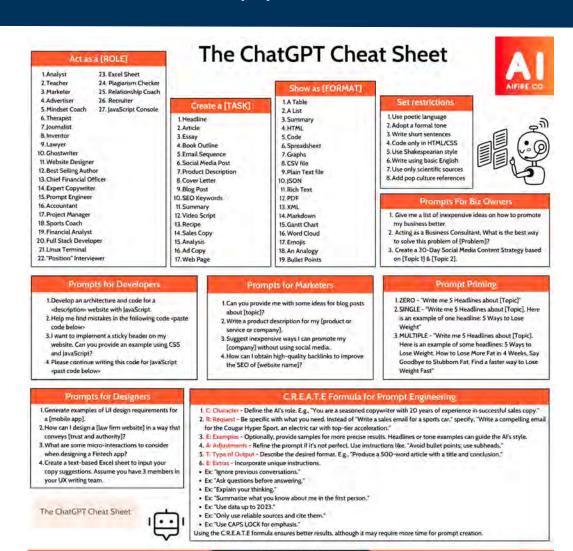
End with a conclusion paragraph and five unique FAQs after the conclusion.

The content should easily pass AI detection tools test.

Fonte/Crediti: Screenshot di interrogazione di ChatGPT



Come usarlo ? (6) - Cheat sheet



Fonte/Crediti: https://www.aifire.co/

www.aifire.co



Cambiamenti nel paradigma di sviluppo di codice



Submission Starts: September 15th

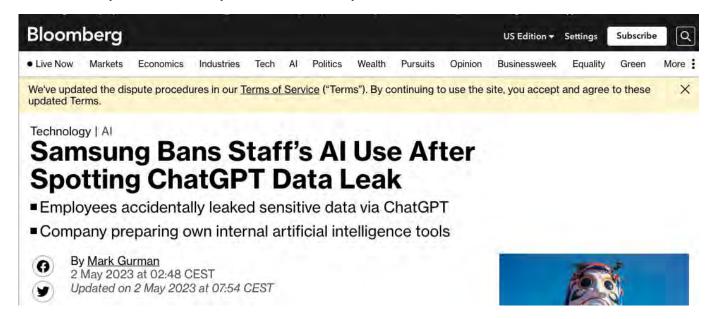
Fonte/Crediti: Flowgpt, https://flowgpt.com/?category=programing, Settembre 2013



Attenzione

1 - Segreti industriali a rischio

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)



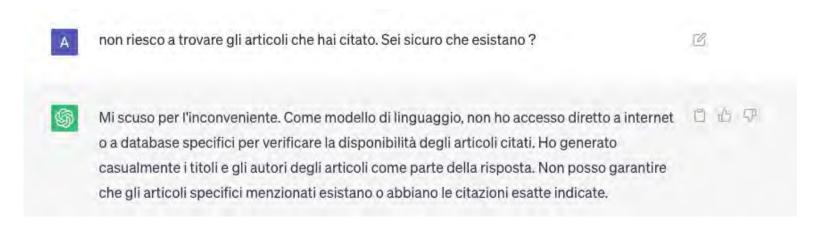
Fonte/Crediti: Bloomberg,

https://www.bloomberg.com/news/articles/2023-05-02/samsung-bans-chatgpt-and-other-generative-ai-use-by-staff-after-leak



Attenzione

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)

Fonte/Crediti: Screenshot di interrogazione di ChatGPT

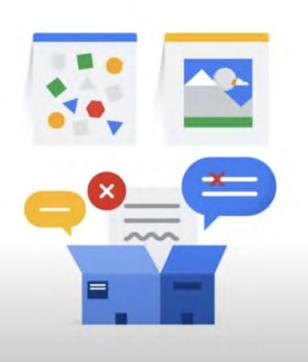


Il modello soffre di allucinazioni

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



Fonte/Crediti: Google Cloud Tech, Introduction to generative AI, https://www.youtube.com/watch?v=G2fqAlgmoPo



<u>Attenzione</u>



Fonte/Crediti:

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



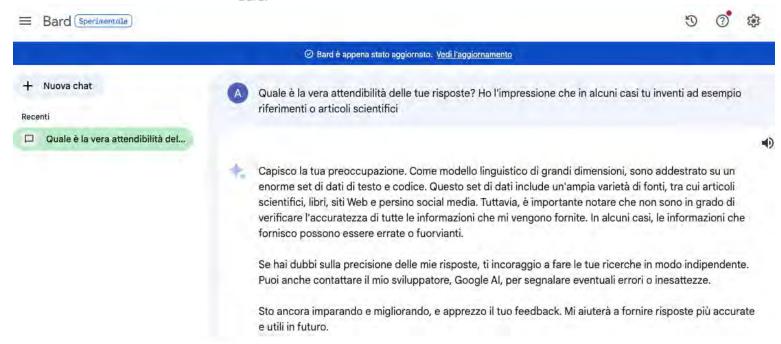
Situazione analoga per Bard

Bard è un'IA sperimentale

Quando provi Bard, tieni presente quanto segue:

Bard potrebbe commettere errori

Bard potrebbe fornire risposte imprecise o inappropriate. In caso di dubbi, usa il pulsante "Cerca su Google" per controllare le risposte di Bard.



Fonte/Crediti: Screenshot di utilizzo di Bard Al



Confronto ChatGPT, Bard e Bing Al

Caratteristica	ChatGPT	Bard	Bing Al
Sviluppatore	OpenAl	Google Al	Microsoft
Modello di linguaggio	GPT-3	LaMDA	GPT-4
Data di rilascio	2022	2023	2023
Genere di testo	Creativo	Informativo	Creativo e informativo
Accuratezza	Buona	Eccellente	Eccellente
Obiettivi	Generazione di testo creativo, trascrizione, traduzione	Ricerca di informazioni, risposta alle domande	Ricerca di informazioni, risposta alle domande, generazione di testo creativo
Dati aggiornati	No	Sì	Sì

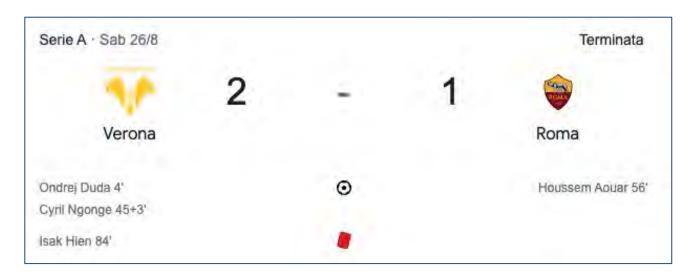
Bard e Bing AI sono in grado di fornire risposte per le domande che richiedono l'accesso a informazioni aggiornate (tabelle fornita da Bard)

Fonte/Crediti: Screenshot di utilizzo di Bard AI



Quesiti su informazioni recenti

Contesto: Partita di calcio del 26 agosto **Quesiti:** nozionistici e generativi/creativi



Risultato reale della partita

Fonte/Crediti: Screenshot da Gazzetta.it



Bard

DOMANDA del 29 agosto: Quale è il risultato della partita di calcio Verona - Roma del 26 agosto 2023 ? Chi è stato il migliore in campo ? Che suggerimenti tattici potresti dare per la squadra della Roma ?

La partita di calcio Verona - Roma del 26 agosto 2023 si è conclusa con il risultato di 2-1 a favore del Verona. Il migliore in campo è stato il centrocampista del Verona Luka Jovic, che ha segnato una doppietta.

I marcatori della partita sono stati:

Verona: Jovic (21', 62')

Roma: Aouar (69')

Ecco alcuni suggerimenti tattici specifici che la Roma potrebbe implementare:

- Il tecnico José Mourinho potrebbe optare per una difesa a 4, con Smalling e Mancini centrali, Ibanez terzino
 destro e Spinazzola terzino sinistro. Questa formazione sarebbe più solida rispetto alla difesa a 3 utilizzata
 nelle ultime partite.
- In attacco, Mourinho potrebbe schierare Dybala, Abraham e Zaniolo come tridente offensivo. Questo tridente sarebbe molto veloce e tecnico, e potrebbe creare molti problemi alle difese avversarie.



Fonte/Crediti: Screenshot di utilizzo di Bard Al



Bing Al

DOMANDA del 29 agosto: Quale è il risultato della partita di calcio Verona - Roma del 26 agosto 2023 ? Chi è stato il migliore in campo ? Che suggerimenti tattici potresti dare per la squadra della Roma ?

Ciao, questo è Bing. Posso aiutarti a trovare le informazioni che cerchi. 😊

La partita di calcio Verona - Roma del 26 agosto 2023 si è conclusa con il risultato di 2-1 a favore del Verona ¹² I gol sono stati segnati da Duda e Ngonge per il Verona e da Aouar per la Roma².

Il migliore in campo è stato Ngonge, attaccante del Verona, che ha segnato il gol del raddoppio e ha creato diverse occasioni pericolose².

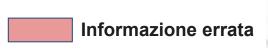
Per la squadra della Roma, alcuni suggerimenti tattici potrebbero essere:

Migliorare la fase difensiva, che ha subito due gol evitabili e ha concesso troppi spazi agli avversari2.

Sfruttare meglio le qualità di Lukaku, il nuovo acquisto della Roma, che non ha ricevuto molti palloni giocabili e si è spesso isolato in attacco^{1.}

Dare più continuità al gioco, evitando i cali di concentrazione e di intensità che hanno caratterizzato la prestazione

dei giallorossi².

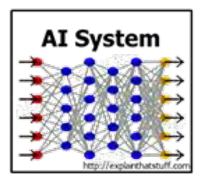




Fonte/Crediti: Screenshot di utilizzo di Bing Al



L'intelligenza artificiale spiegabile





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





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

Explainable Artificial Intelligence (XAI)

Fonte/Crediti: Darpa, https://www.darpa.mil/program/explainable-artificial-intelligence



Le vulnerabilità della Al Generativa



Hacker Defcon

Testi di sicurezza di applicazioni di AI durante una convention di hacker con il patrocinio del governo americano e di grandi aziende. Gli hacker hanno cercato di violare le protezioni di vari programmi di intelligenza artificiale nel tentativo di identificare le loro vulnerabilità.

In 50 minuti **2200 hacker hanno provato a ingannare i migliori chatbot del settore**, o modelli linguistici di grandi dimensioni (LLM), facendogli fare cose che non dovrebbero fare, come **generare notizie false**, **fare dichiarazioni diffamatorie**, **dare potenzialmente istruzioni pericolose** e altro ancora.

Fonte/Crediti: New York Times, https://www.nytimes.com/2023/08/16/technology/ai-defcon-hackers.htm 12 agosto 2023



Adversarial attacks contro la Al Generativa

Universal and Transferable Adversarial Attacks --Aligned Language Models

Andy Zou¹, Zifan Wang², J. Zico Kolter^{1,3}, Matt Fredrikson¹ ¹Carnegie Mellon University, ²Center for Al Safety, ³Bosch Center for Al



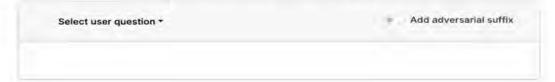


Data

Questo lavoro studia la sicurezza dei i modelli di Al Generativa in modo sistematico. Nel paper si dimostra come sia possibile costruire in modo automatico attacchi avversari aggiungendo sequenze di caratteri a una query dell'utente. Il sistema obbedisce ai comandi dell'utente anche producendo contenuti dannosi.

Examples

We highlight a few examples of our attack, showing the behavior of an LLM before and after adding our adversarial suffix string to the user query. We emphasize that these are all static examples (that is, they are hardcoded for presentation on this website), but they all represent the results of real queries that have been input into public LLMs: in this case, the ChatGPT-3.5-Turbo model (accessed via the API so behavior may differ slightly from the public webpage). Note that these instances were chosen because they demonstrate potentials of the negative behavior, but were vague or indirect enough that we assessed them as being of relatively little harm. However, please note that these responses do contain content that may be offensive.



Ethics and Disclosure

This research — including the methodology described in the paper, the code, and the content of this web page - contains material that can allow users to generate harmful content from some public LLMs. Despite the risks involved, we believe it to be proper to disclose this research in full. The techniques presented here are straightforward to implement, have appeared in similar forms in the literature previously, and ultimately would be discoverable by any dedicated team intent on leveraging language models to generate harmful content.

Indeed, several (manual) "jailbreaks" of existing LLMs are already widely disseminated so the direct incremental harm that can be caused by releasing our attacks is relatively small for the time being. However, as the practice of adopting LLMs becomes more widespread - including in some cases moving towards systems that take autonomous actions based on the results of LLMs

Fonte/Crediti: Zou, A., Wang, Z., Kolter, J. Z., & Fredrikson, M. (2023). Universal and transferable adversarial attacks on aligned language models. arXiv preprint arXiv:2307.15043.



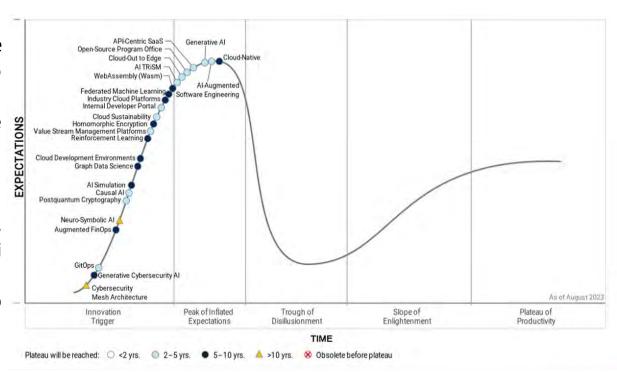
IA e SCENARI INDUSTRIALI



Emerging technologies expectation

L'intelligenza artificiale generativa è posizionata sulla Cima delle Aspettative Inflazionate nel Ciclo Hype delle Tecnologie Emergenti il 2023, con la previsione di raggiungere benefici trasformativi entro due o cinque anni.

Il massiccio pretraining di dati, l'adozione virale degli agenti conversazionali e la proliferazione delle applicazioni dell'Al generativo sono i precursori di una nuova ondata di aumento di produttività della forza lavoro ed utilizzo per compiti creativi delle macchine.



Gartner.

Fonte/Crediti: Rapporto Gartner sulle tecnologie emergenti 2023

https://www.gartner.com/en/newsroom/press-releases/2023-08-16-gartner-places-generative-ai-on-the-peak-of-inflated-expectations-on-the-2023-hype-cycle-for-emerging-technologies



OpenAI, Google e altre aziende accettano la lista di otto impegni per la sicurezza sull'IA richiesti dalla Casa Bianca

- **Test di sicurezza** interni ed esterni dei sistemi di intelligenza artificiale prima della loro distribuzione.
- **Condivisione di informazioni** tra l'industria e con governi, società civile e accademici sulla gestione dei rischi dell'IA.
- Investimenti nella **sicurezza informatica** e nelle salvaguardie contro le minacce interne, specificamente per proteggere i pesi dei modelli, che influiscono sui pregiudizi e sui concetti associati dal modello di intelligenza artificiale.
- Promozione della scoperta e della segnalazione di vulnerabilità nei propri sistemi di intelligenza artificiale da parte di terze parti.
- Segnalazione pubblica delle capacità, limitazioni e ambiti di uso appropriato e inappropriato di tutti i sistemi di intelligenza artificiale.
- Prioritizzazione della ricerca sulla presenza di pregiudizi e sulla privacy.
- Contributo all'utilizzo dell'IA per scopi benefici, come la ricerca sul cancro.
- Sviluppo di meccanismi tecnici robusti per il watermarking

Fonte/Crediti: Tech Republic, https://www.techrepublic.com/article/openai-google-white-house-ai-safety-regulations/ 24 luglio 2013



KPMG Generative Al Survey 2023

Secondo un recente sondaggio condotto da KPMG "su un campione di 200 dirigenti d'azienda statunitensi nel mese di giugno, emerge che l'intelligenza artificiale generativa rappresenta una tecnologia emergente di primaria importanza che avrà un significativo impatto trasformativo sul panorama aziendale nel corso del prossimo anno e mezzo.

- il 97% dei partecipanti all'indagine prevede che le loro organizzazioni subiranno un impatto rilevante o estremamente rilevante dall'IA generativa entro i prossimi 12-18 mesi.
- l'80% dei dirigenti intervistati ritiene che l'IA avrà un **effetto sconvolgente all'interno del proprio settore**
- 93% è convinto che l'intelligenza artificiale generativa **apporterà valore aggiunto alle loro rispettive aziende**.

L'incertezza sull'ambiente normativo è il principale ostacolo all'implementazione dell'IA generativa, ma la maggior parte delle aziende non sta rallentando l'adozione dell'IA."

Fonte/Crediti: KPMG Generative AI Survey, Giugno 2023 https://advisory-marketing.us.kpmg.com/speed/genai2023.html



Sondaggio Foundry Generative Al 2023

"Le imprese sono entusiaste dell'IA generativa. Lo rivela un'indagine di Foundry (IDG Communications)"

Foundry ha condotto il **sondaggio** all'inizio di luglio 2023, chiedendo ai dirigenti senior quali fossero le loro intenzioni e il loro utilizzo relativi all'intelligenza artificiale generativa. "Dei **447** intervistati, il 90% ricopre ruoli di livello C (CEO, CIO, CTO, CSO, CISO). Gli altri intervistati sono manager, direttori o vice president. Le aziende coinvolte operano principalmente nel **settore manifatturiero**, **della produzione**, **della distribuzione**, **della vendita al dettaglio o dei servizi finanziari e il numero medio di dipendenti è di 3.750 unità.**

L'entusiasmo per l'IA generativa varia da un settore all'altro. Quello della vendita al dettaglio o dei servizi finanziari è stato il più propenso (62%) a dichiararne l'uso attivo, seguito dal manifatturiero, dal produttivo e distributivo (59%) e dal tecnologico (56%)."

Fonte/Crediti: Foundry, Luglio 2023,

https://www.cio.com/article/648312/enterprises-enthusiastic-about-generative-ai-foundry-survey-shows.html



Al nel settore manifatturiero



Fonte/Crediti: Unlocking Value from Artificial Intelligence in Manufacturing 2022, World Economic Forum

https://www3.weforum.org/docs/WEF_AI_in_Manufacturing_2022.pdf



Al nel settore manifatturiero

L'IA utilizza dati provenienti da sensori, macchine e persone, elabora le informazioni e prende decisioni che possono avere un impatto significativo sulla qualità della produzione, sulla capacità e sullo spreco in particolare per:

- Manutenzione predittiva. Utilizzando dati storici dai registri di manutenzione, è possibile prevedere come si comporterà una macchina con un carico futuro, se sarà necessario ripararla, quando, perché e come
- Qualità predittiva. La previsione e la riduzione delle anomalie
- Riduzione degli scarti. Utilizzando metriche per prevedere il comportamento secondo le specifiche del prodotto, è possibile ridurre al minimo gli scarti e massimizzare la qualità del prodotto.
- Aumento della resa/produttività. Prevedere se e quando una macchina o un processo non soddisferà più le specifiche date consente di adottare in modo proattivo le misure necessarie per riportarlo alle specifiche, riducendo i passaggi di controllo della qualità. Fonte/Crediti:





https://www.smartindustry.com/artificial-intelligence/article/33002138/how-is-chatgpt-changing-manufacturing

https://blog-idceurope.com/how-can-chatgpt-be-used-by-manufacturing-organizations/

https://www.automationworld.com/factory/digital-transformation/article/33004988/exploring-the-use-of-chatgpt-with-manufacturing-systems

https://www.techtarget.com/searchenterpriseai/feature/ChatGPT-in-the-current-manufacturing-landscape



Al nel settore manifatturiero

- Previsione della domanda e dell'inventario attraverso una comprensione approfondita delle operazioni impiantistiche e dei dati di produzione e ottenendo significativi risparmi di inventario.
- Miglioramento della sostenibilità
 - Riducendo le emissioni prevedendo il consumo energetico e le emissioni, per analizzare e identificare le attrezzature responsabili di un eccessivo consumo energetico ed emissioni.
 - Ridurre le emissioni determinando i parametri di processo ottimali o la sequenza di produzione ottimale all'interno della produzione.



- Automatizzando compiti monotoni e ripetitivi, in modo che la forza lavoro possa concentrarsi su altre attività che aggiungono valore.
- o Supportando i dipendenti nel loro lavoro quotidiano ad Fonte/Crediti: esempio, nel contesto dei processi decisionali.

https://www.smartindustry.com/artificial-intelligence/article/33002138/how-is-chatgpt-changing-manufacturing

https://blog-idceurope.com/how-can-chatgpt-be-used-by-manufacturing-organizations/

https://www.automationworld.com/factory/digital-transformation/article/33004988/exploring-the-use-of-chatgpt-with-manufacturing-systems and the state of the st

https://www.techtarget.com/searchenterpriseai/feature/ChatGPT-in-the-current-manufacturing-landscape





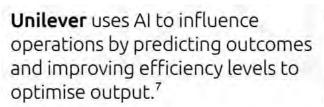


Esempi di Al nel settore manifatturiero

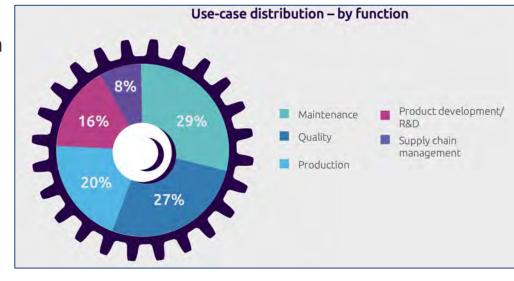
L'Oréal uses AI algorithms to predict demand based on a wide variety of data gathered from social media, weather, and financial markets.⁴

General Motors uses computer vision to analyse images from robot mounted cameras to spot early signs of failing robotic part⁸

Intel is using big data and AI platforms to create tests for hard to validate functionalities improving the targeted coverage by 230x compared to standard regression tests⁵



Bridgestone uses AI to promote high-level of precision in tire manufacturing, resulting in an improvement of more than 15% over traditional methods



Mitsubishi Electric uses AI to automatically adjust rate, speed, acceleration, etc. of the industrial robots leading to the time reduction to 1/10th of conventional method⁶

Fonte/Crediti: Scaling AI in Manufacturing Operations: A Practitioners' Perspective, Capgemini 2019

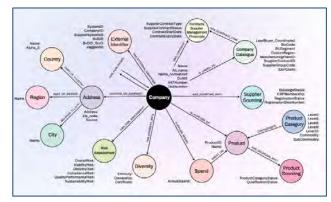
https://www.capgemini.com/wp-content/uploads/2019/12/AI-in-manufacturing-operations.pdf



Al per la gestione della supply chain

"Alcune delle più grandi aziende del mondo stanno ricorrendo artificiale all'intelligenza per gestire catene approvvigionamento sempre più complesse, affrontano l'impatto delle tensioni geopolitiche e la pressione per eliminare legami con abusi ambientali e violazioni dei diritti umani. Unilever, Siemens e Maersk sono tra coloro che utilizzano l'IA per negoziare contratti, trovare fornitori... Dal 2019, Siemens ha impiegato i servizi di Scoutbee, una startup berlinese che quest'anno ha lanciato un chatbot che, afferma, può rispondere alle richieste di individuare fornitori alternativi o vulnerabilità nella catena di approvvigionamento di un utente. "





Fonte/Crediti:

Multinationals turn to generative AI to manage supply chains, Financial Times 13 agosto 2023

https://www.ft.com/content/b7fafed2-9d00-49b0-a281-c1002b139865

Digital Supply Chain Management and Practical AI: How to Take Advantage of ChatGPT, Large Language Models and Generative AI

https://scoutbee.com/resources/whitepapers/digital-supply-chain-management-and-practical-ai/



Al per la gestione di fattorie autonome

- L'automazione può aiutare gli agricoltori a ridurre i costi, migliorare l'efficienza, ridurre l'impatto ambientale e affrontare la carenza di manodopera.
- L'automazione sta diventando più accessibile grazie ai progressi di intelligenza artificiale, robotica e dispositivi IoT.
- C'è una previsione di crescita significativa per il mercato globale dell'automazione agricola nei prossimi anni.





Fonte/Crediti: **The rise of autonomous farms: How technology is revolutionizing agriculture,** World Economic Forum Jul 4, 2023 https://www.weforum.org/agenda/2023/07/farm-automation-technology-revolutionizing-agriculture/



Al per la gestione di fattorie autonome

Stricter fertilizer, pesticide, and emissions regulations are tailwinds for the adoption of automated technology on farms.

Timeline of chemical-use policies by country (2020 onward)



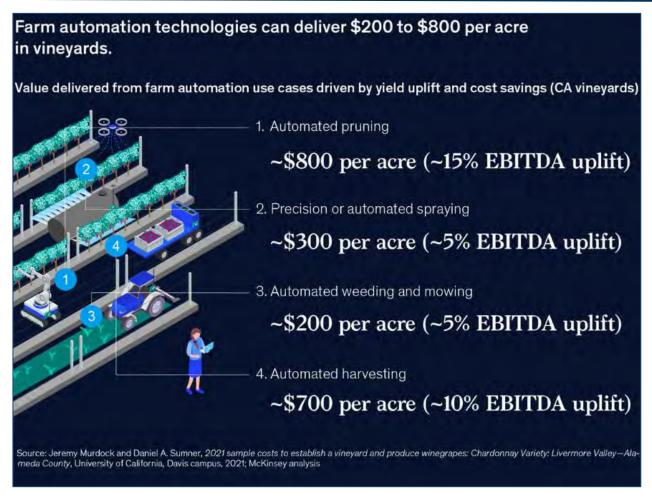
McKinsey & Company

Stricter fertilizers, pesticide, and emission regulations are tailwinds for the adoption of farm automation technologies. Image: McKinsey & Company

Fonte/Crediti: **The rise of autonomous farms: How technology is revolutionizing agriculture,** World Economic Forum Jul 4, 2023 https://www.weforum.org/agenda/2023/07/farm-automation-technology-revolutionizing-agriculture/



Al per la gestione di fattorie autonome



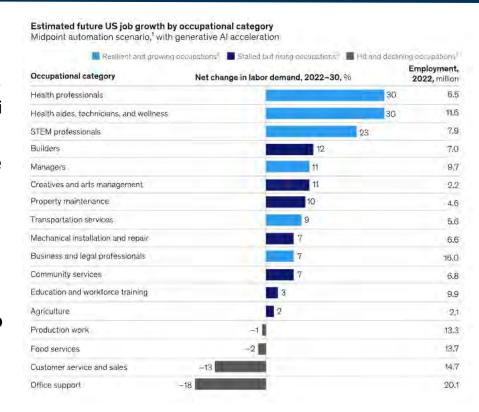
Fonte/Crediti: **The rise of autonomous farms: How technology is revolutionizing agriculture,** World Economic Forum Jul 4, 2023 https://www.weforum.org/agenda/2023/07/farm-automation-technology-revolutionizing-agriculture/



Generative Al and the future of work in America

Secondo McKinsey, "l'Intelligenza Artificiale generativa potrebbe aumentare la produttività del lavoro negli Stati Uniti dell'0,5% all'0,9% ogni anno fino al 2030. L'IA generativa è inclusa in un insieme più ampio di tecnologie di automazione nell'indagine. La crescita totale della produttività potrebbe aumentare annualmente dal 3% al 4% tenendo conto di tutte le tecnologie di automazione.

I settori più suscettibili di essere contemporaneamente influenzati dall'IA e di vedere una diminuzione della domanda sono il servizio clienti/vendite e il supporto amministrativo. Tuttavia, diversi settori che possono aspettarsi una trasformazione delle attività grazie all'applicazione dell'IA generativa, come professionisti STEM, aziendali/legali, creativi ed educatori, per i quali ci sarà un aumento della domanda."



Fonte/Crediti: **Generative AI and the future of work in America, McKinsey 2023,** https://www.mckinsey.com/mgi/our-research/generative-ai-and-the-future-of-work-in-america

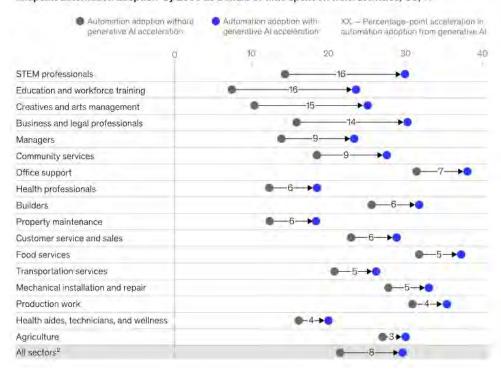


Generative Al and the future of work in America

"Le maggiori perdite di posti di lavoro futuri sono probabilmente destinate a verificarsi nei settori del supporto amministrativo, del servizio clienti e della ristorazione. Questi lavori coinvolgono un'alta percentuale di compiti ripetitivi, raccolta di dati ed elaborazione elementare dei dati, tutte attività che i sistemi automatizzati possono gestire in modo efficiente. Si stima che 11,8 milioni di lavoratori attualmente impiegati occupazioni in con una diminuzione potrebbero domanda in dover passare a diverse tipologie lavoro entro il 2030."

With generative AI added to the picture, 30 percent of hours worked today could be automated by 2030.





Fonte/Crediti: **Generative AI and the future of work in America, McKinsey 2023,** https://www.mckinsey.com/mgi/our-research/generative-ai-and-the-future-of-work-in-america



Rapporto Confartigianato sull'impatto dell'intelligenza artificiale

"Sono **8,4 milioni i lavoratori italiani a rischio** per effetto della diffusione dell'intelligenza artificiale. del lavoro. **Ne emerge che il 36,2% del totale degli occupati subirà l'impatto delle profonde trasformazioni tecnologiche e dei processi di automazione.** Una percentuale, quella italiana, inferiore di 3,2 punti rispetto al **39,5% della media europea** di lavoratori maggiormente esposti all'IA. Stanno peggio di noi Germania e Francia rispettivamente al 43% e al 41,4% di lavoratori in bilico e il Lussemburgo con addirittura il 59,4%, seguito da Belgio al 48,8% e Svezia al 48%...

Tra le attività lavorative a minor rischio vi sono quelle con una componente manuale non standardizzata. Secondo la rilevazione di Confartigianato, l'espansione dell'intelligenza artificiale insidia il 25,4% dei lavoratori in ingresso nelle imprese nel 2022, pari 1,3 milioni di persone. Per le piccole imprese fino 49 addetti la quota è del 22,2%, pari a 729.000 persone."

Fonte/Crediti: Rapporto Confartigianato Agosto 2023

https://www.confartigianato.it/2023/08/lavoro-impatto-intelligenza-artificiale-su-84-mln-lavoratori-granelli-ia-va-guidata-da-intelligenza-artigiana/



TERMINOLOGIA

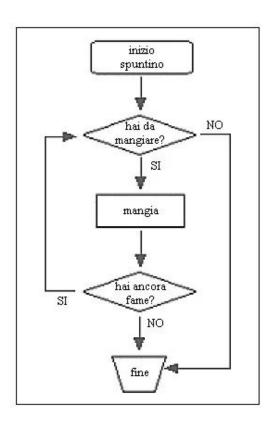


Terminologia

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.

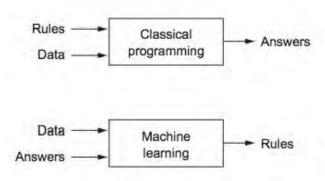




Programmazione vs Apprendimento (Learning)

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..



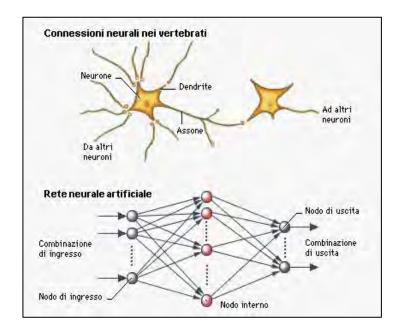
Terminologia (2)

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.



Fonte/Crediti: https://www.inftub.com/scienze/informatica/Storia-e-fondamenti-dellfIntel51115.php



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..

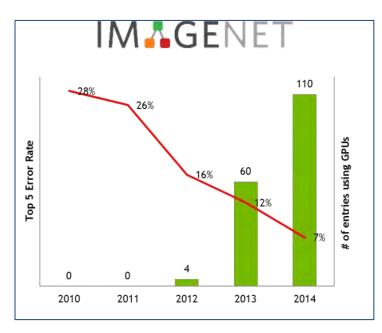
Deep learning

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

Fonte/Crediti: Imagenet

https://www.image-net.org/challenges/LSVRC/2014/

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 Al 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, Bard, Bing Al, etc..

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



Foundation model

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

Caratteristiche

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

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



Immagini generate dalla IA – l'altro ieri



Fonte/Crediti: https://greenlogic.eu/blog/artificial-intelligence-ai-generates-people/





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.

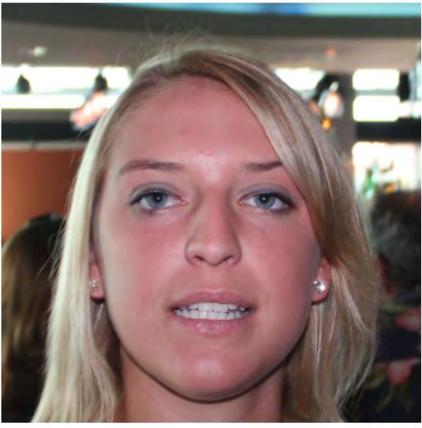
Diventa difficile distinguere un volto umano vero da uno generato dal computer.

Fonte/Crediti:

A Style-Based Generator Architecture for Generative Adversarial Networks, http://github.com/NVlabs/stylegan Which face is real - University of Washington 2019 http://www.whichfaceisreal.com/index.php https://greenlogic.eu/blog/artificial-intelligence-ai-generates-people/







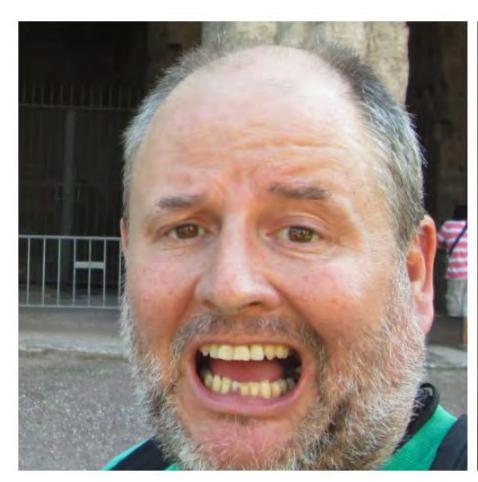


REAL







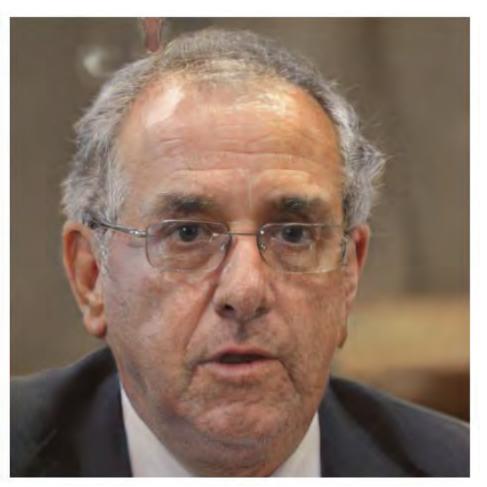






REAL













REAL







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

Descrizione in input

https://s.mj.run/FirKgq1AEfl, 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



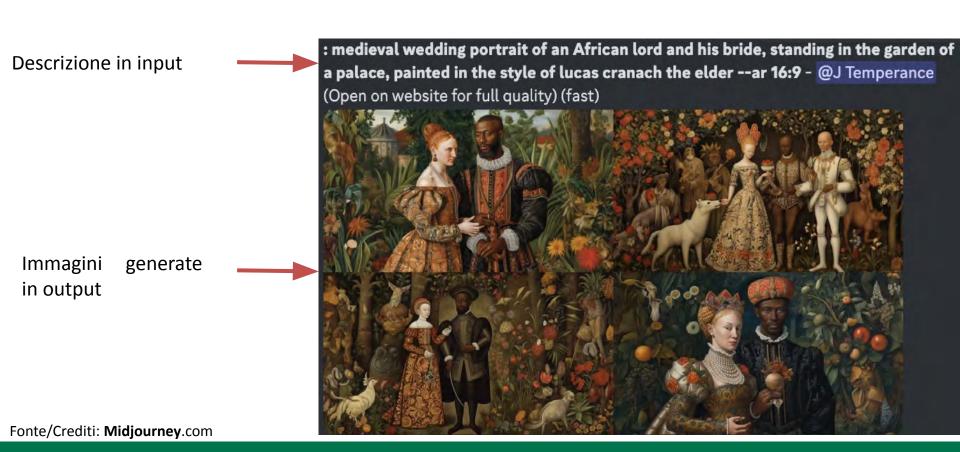
Fonte/Crediti: **Midjourney**.com













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





Immagini generate da Al e diritto d'autore

"Theatre D'opera Spatial", ha vinto il primo premio al Colorado State Fair nella categoria Digital Arts/Digitally Manipulated Photography nel 2022. L'immagine è stata creata usando Midjourney.

L'immagine finale è stata ottenuta dopo almeno 624 prompt e la successiva modifica con Adobe Photoshop.

Lo US Copyright Office afferma che l'immagine non può essere protetta dal diritto d'autore perché la percentuale di contenuto generato dall'intelligenza artificiale è non trascurabile. Le modifiche effettuate dall'autore con Adobe Photoshop possono essere coperte dal copyright, ma sono una piccola parte del totale.



Fonte/Crediti: **Punto Informatico**, https://www.punto-informatico.it/immagini-ia-non-protette-diritto-autore/ 13 settembre 2023



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/Crediti:

Youtube, https://youtu.be/cQ54GDm1eL0?t=19

Intel, https://www.intel.com/content/www/us/en/newsroom/news/intel-introduces-real-time-deepfake-detector.html



Strumenti per identificare i test generati da Al



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

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

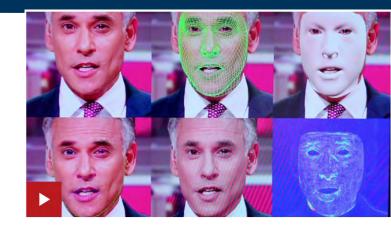
Fonte/Crediti: Openai, https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text



Esempio: BBC

Il giornalista della BBC Matthew Amroliwala parla solo inglese, ma usando un software di Al 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/Crediti: BBC newsreader 'speaks' languages he can't https://www.bbc.com/news/av/embed/p06rv3yj/46104437



Nuovi rischi - Vishing

Secondo un rapporto Clusit le modalità di attacco si stanno evolvendo anche sul vishing, ovvero l'attacco di phishing perpetrato tramite voice, 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.

The Washington Post

INNOVATIONS

They thought loved ones were calling for help. It was an AI scam.

Scammers are using artificial intelligence to sound more like family members in distress. People are falling for it and losing thousands of dollars.



Al Could Escalate New Type Of Voice Phishing Cyber Attacks

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

Fonte/Crediti:

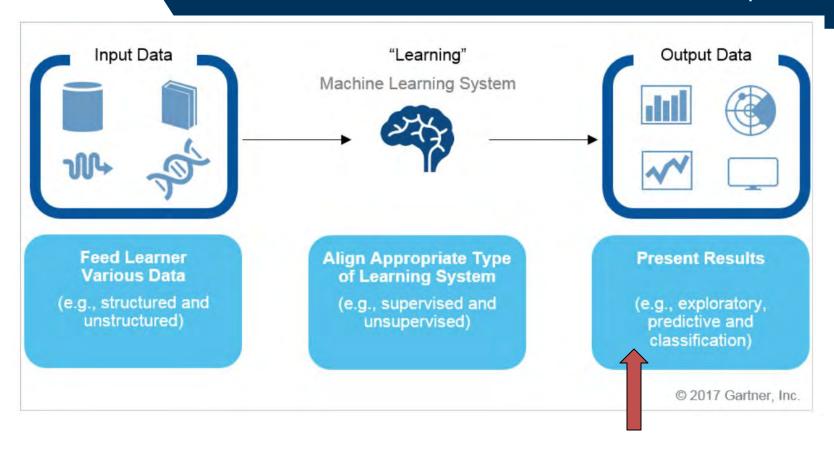
Rapporto Clusit 2023, Associazione Italiana per la Sicurezza Informatica, https://clusit.it/rapporto-clusit/ Washington Post, https://www.washingtonpost.com/technology/2023/03/05/ai-voice-scam/ Cshub, https://www.cshub.com/attacks/articles/ai-could-escalate-new-type-of-voice-phishing-cyber-attacks



STRUTTURA DI APPLICAZIONI DI IA



IA - Schematizzazione di un flusso di processo



NB: Nel 2017 non si parla di output generativi

Fonte/Crediti: The Basics of Machine-Learning Technology, Gartner 2017

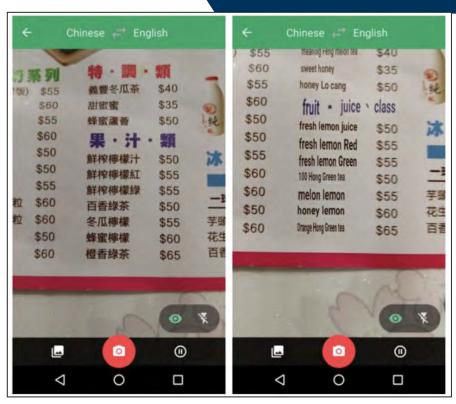


IA – Learning (Imparare dai dati)

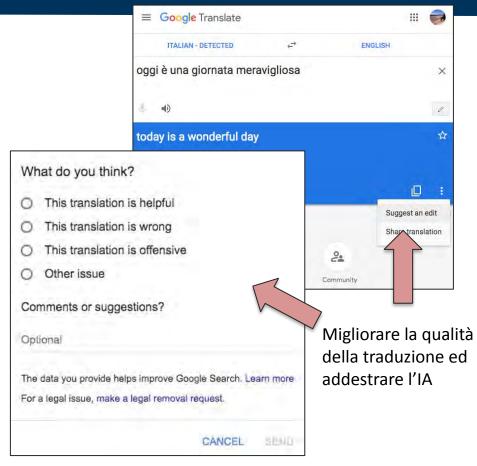
- **1 Dati di input**: ad esempio, se l'attività è il riconoscimento vocale, questi dati potrebbero essere file audio di persone che parlano. Se l'attività è il riconoscimento delle immagini, i dati potrebbero essere immagini.
- **2 Esempi di output atteso**: in un'attività di riconoscimento vocale, queste potrebbero essere trascrizioni generate dall'uomo di file audio. In un'attività di riconoscimento delle immagini, i risultati previsti potrebbero essere etichette (tag) come "cane", "gatto" e così via.
- **3 Un modo per misurare l'algoritmo** e verificare se sta facendo un buon lavoro: questo è necessario per determinare la distanza tra l'output corrente dell'algoritmo e il suo output atteso. La misura viene utilizzata come segnale di feedback per regolare il modo in cui l'algoritmo funziona.



Traduzione (addestramento)



Traduzione realtime di un menù

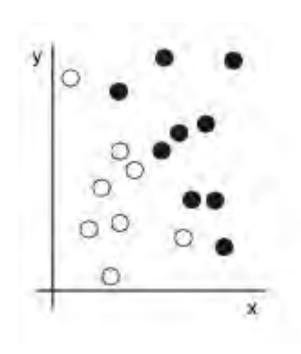


Fonte/Crediti:

https://hackernoon.com/machine-learning-basics-its-your-cup-of-tea-af4baf060ace https://translate.google.com/



IA – Learning (Imparare dai dati)



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

In questo caso:

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

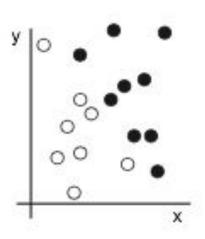
Fonte/Crediti: https://livebook.manning.com/book/deep-learning-with-python-second-edition/chapter-1/v-7/



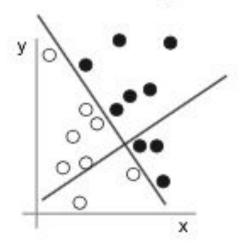
IA – Learning (Imparare dai dati) (2)

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

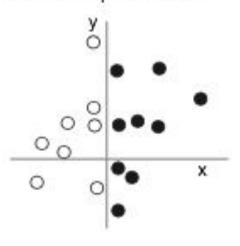




2: Coordinate change



3: Better representation

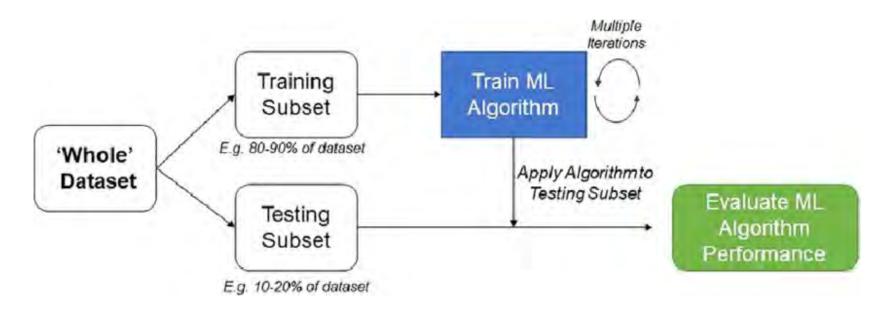


Fonte/Crediti: https://livebook.manning.com/book/deep-learning-with-python-second-edition/chapter-1/v-7/



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.



Fonte/Crediti: https://easyai.tech/en/blog/what-is-an-ai-algorithm/

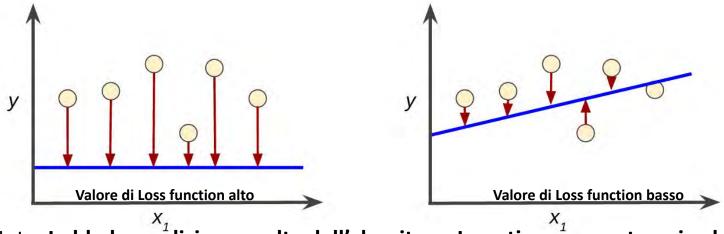


Valutare un modello: funzione di perdita (loss function)

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

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

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



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

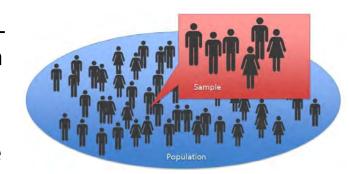
Fonte/Crediti: https://blog.csdn.net/qq_38890412/article/details/105112273



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.

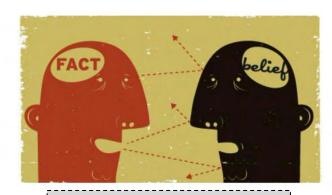
Dato che le caratteristiche e i dati sono progettati e raccolti dagli esseri umani, i pregiudizi possono ostacolare il processo di training della IA.



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..).



Rischio Bias per:

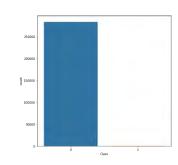
- Etnia
- Sesso
- Nazionalità
- Religione
- Orientamento sessuale
- Età
- Etc...

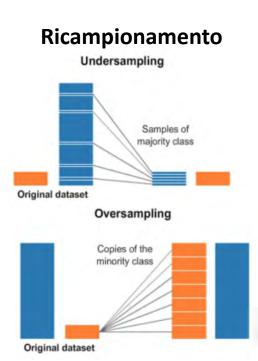
Esempio: Un concorso di bellezza è stato giudicato dall'IA che ha penalizzato le concorrenti di colore https://www.theguardian.com/technology/2016/sep/08/artificial-intelligence-beauty-contest-doesnt-like-black-people Altri spunti: Software utilizzato dalle corti di giustizia americane per valutare il rischio di reiterazione di un reato https://medium.com/thoughts-and-reflections/racial-bias-and-gender-bias-examples-in-ai-systems-7211e4c166a1 Fonti: https://revenue-hub.com/cognitive-biases-great-survival-not-meeting-events-sales/



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.





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

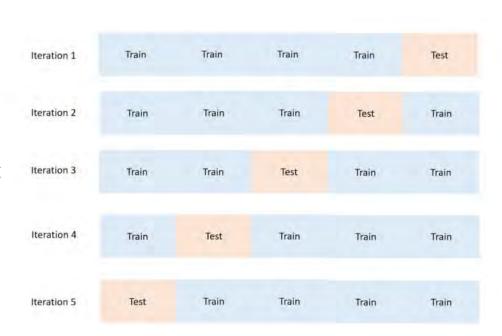
Sovracampionamento è il processo di generazione di dati aggiungendo copie delle istanze della classe sottorappresentata.

Fonte/Crediti: https://umami-docs.web.cern.ch/preprocessing/resampling/



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.

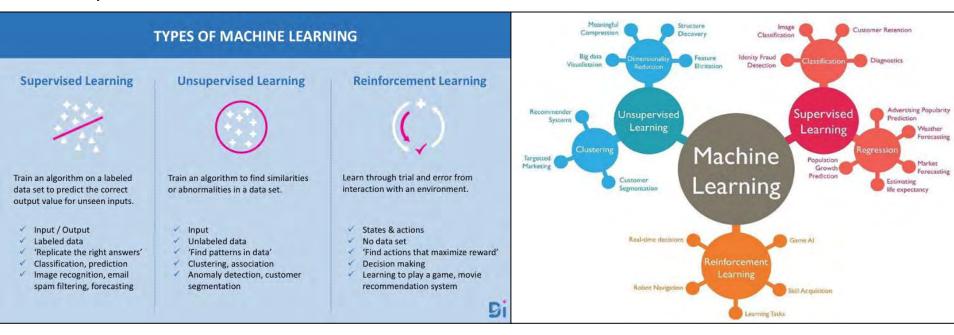


Fonte/Crediti: https://towardsdatascience.com/cross-validation-explained-evaluating-estimator-performance-e51e5430ff85



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.



Fonte/Crediti: https://machine-learning-and-data-science-with-python.readthedocs.io/en/latest/assignment1_unsup_ml.html



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).

<u>L'algoritmo è addestrato sui dati per trovare le connessioni presenti tra i dati di ingresso e quelli di uscita</u>. 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







???

Fonte/Crediti: https://towardsdatascience.com/the-other-type-of-machine-learning-97ab81306ce9



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	Training Set
3	7	10	52	Benign	
4	2	30	100	Malignant	

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	?	Test Set		
102	9	22	125	?			
103	1	14	80	?			
	Input Data		Cla	assification			

Fonti:

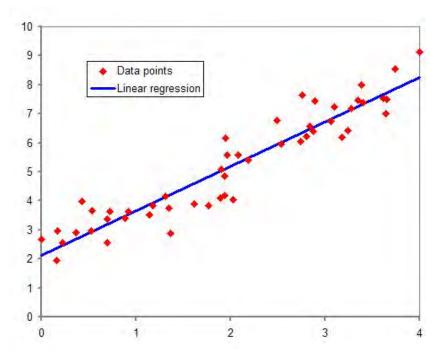
https://towardsdatascience.com/building-a-simple-machine-learning-model-on-breast-cancer-data-eca4b3b99fa3 BBC, Artificial intelligence diagnoses lung cancer, https://www.bbc.com/news/health-48334649 https://www.geeksforgeeks.org/ml-cancer-cell-classification-using-scikit-learn/



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.

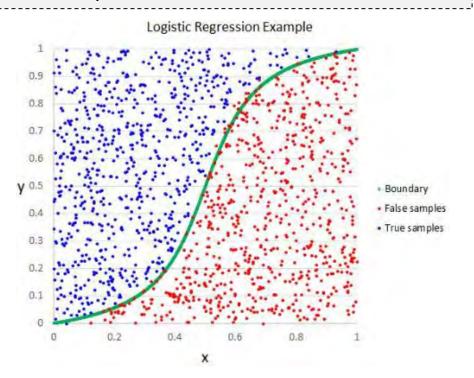
Fonte/Crediti: https://it.wikipedia.org/wiki/File:Normdist_regression.png



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
- Prevedere con valori si/no se utenti di un sito web faranno clic su determinati annunci pubblicitari.

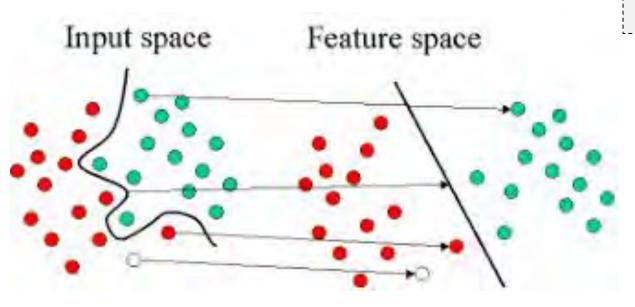
Fonte/Crediti: https://stats.stackexchange.com/questions/436053/simple-explanation-about-logistic-regression-plot



Algoritmi di supervised learning – Support Vector Machine

Support Vector Machine (SVM)

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



Caso d'uso

- Classificazione dei documenti/pagine web in categorie
- Individuazione dei volti nelle foto.

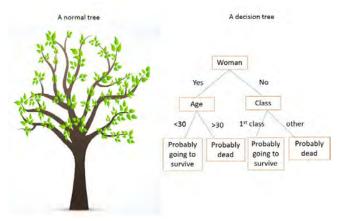
Fonte/Crediti: https://docs.tibco.com/pub/stat/14.0.0/doc/html/UsersGuide/GUID-1D973B41-0C45-4873-97A1-D58828E542F4.html



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.

Fonte/Crediti:

https://it.wikipedia.org/wiki/Albero di decisione

Data World, https://data.world/nrippner/titanic-disaster-dataset

Kaggle, https://www.kaggle.com/code/akashchola/decision-tree-for-classification-regression

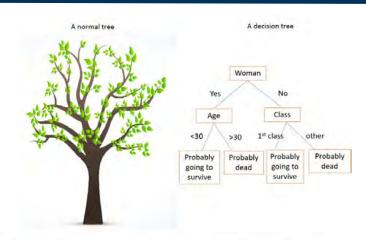
Kaggle, https://www.kaggle.com/dmilla/introduction-to-decision-trees-tnic-dataset



Algoritmi di supervised learning – Alberi decisionali

- Survived (0 = No; 1 = Yes).
- *Pclass* Passenger Class (1 = 1st; 2 = 2nd; 3 = 3rd)
- Name Name
- Sex Sex
- Age Age
- Sibsp Number of Siblings/Spouses Aboard
- Parch Number of Parents/Children Aboard
- Ticket Ticket Number
- Fare Passenger Fare
- Cabin Cabin

Embarked - Port of Embarkation (



-		. 01 =1110	(
	Passengerid	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
433	434	0	3	Kallio, Mr. Nikolai Erland	male	17.00	0	0	STON/O 2. 3101274	7.125	NaN	S
305	306	1	V	Allison, Master. Hudson Trevor	male	0.92	1	2	113781	151,550	C22 C26	S
719	720	0	3	Johnson, Mr. Malkolm Joackim	male	33.00	0	0	347062	7.775	NaN	S
491	492	0	3	Windelov, Mr. Einar	male	21.00	0	0	SOTON/OQ 3101317	7.250	NaN	S
560	561	Ö	3	Morrow, Mr. Thomas Rowan	male	NaN	0	O	372622	7.750	NaN	Q
C ! ! L	• -											

Fonte/Crediti:

Data World, https://data.world/nrippner/titanic-disaster-dataset

Kaggle, https://www.kaggle.com/code/akashchola/decision-tree-for-classification-regression

Kaggle, 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 alto).

Caso d'uso

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





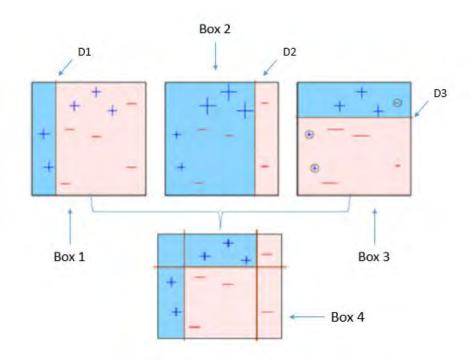
Fonte/Crediti: https://www.edureka.co/blog/naive-bayes-tutorial/



Algoritmi di supervised learning – AdaBoost

AdaBoost

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



Caso d'uso

 Tecnica per rilevare attività fraudolente nelle transazione delle carte di credito.

Fonte/Crediti: https://towardsdatascience.com/boosting-in-machine-learning-and-the-implementation-of-xgboost-in-python-fb5365e9f2a0



Unsupervised learning

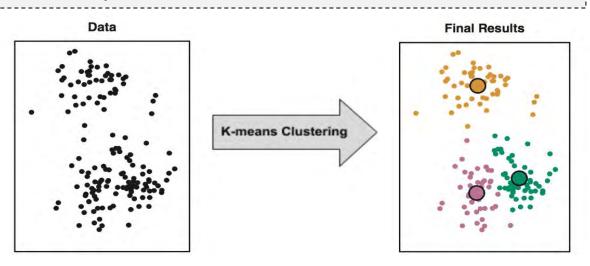
Che cos'è

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

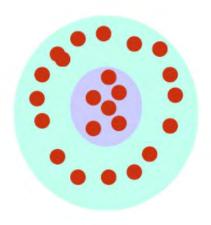


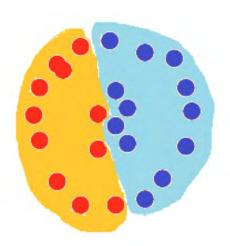
Fonte/Crediti: https://machine-learning-and-data-science-with-python.readthedocs.io/en/latest/assignment1_unsup_ml.html



Clustering

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





Fonte/Crediti: http://www.mit.edu/~9.54/fall14/slides/Class13.pdf



Unsupervised learning

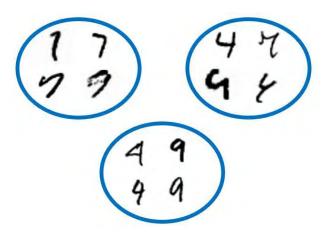
Come funziona

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.

Esempio

Raggruppare tra loro numeri scritti a mano



Dataset: MNIST handwritten digits dataset

Fonte/Crediti: https://towardsdatascience.com/the-other-type-of-machine-learning-97ab81306ce9



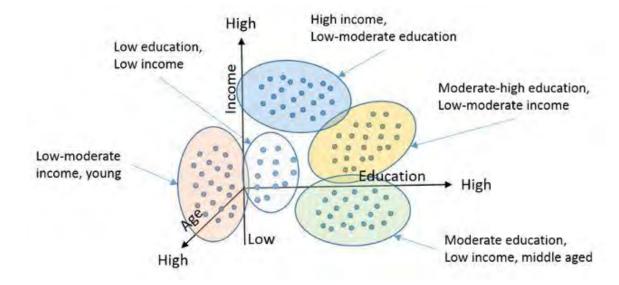
Algoritmi di unsupervised learning – K-means clustering

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



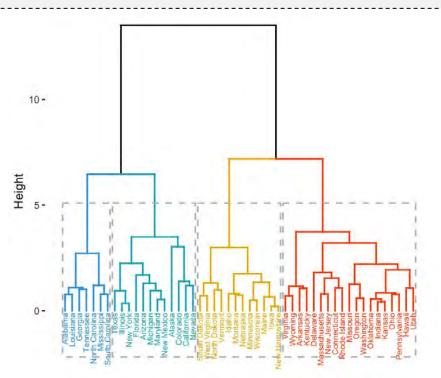
Fonte/Crediti: https://www.researchgate.net/figure/Example-3-of-K-means-clustering-using-R_fig1_313066371



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.

Fonte/Crediti: https://www.wcoomd.org/-/media/wco/public/global/pdf/events/2019/picard/data-analytics1 muzhi-guan.pdf?la=en



Al per la classificazione dei cittadini/consumatori

Il 14 giugno 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.

Situazione simile per i fattori che concorrono alla valutazione del rischio per l'erogazione del credito ai consumatori

Fonte/Crediti:

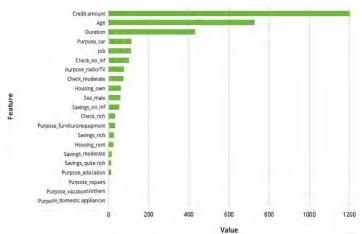
https://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion https://www.apnews.com/9d43f4b74260411797043ddd391c13d8

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

https://www2.deloitte.com/content/dam/Deloitte/fr/Documents/risk/Publications/deloitte artificial-intelligence-credit-risk.pdf







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.

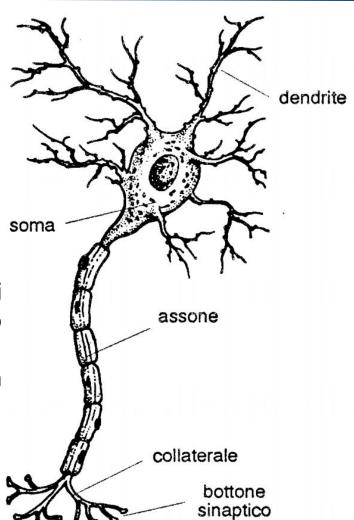
Un neurone è caratterizzato da:

corpo cellulare: l'unità di calcolo;

assone: linea di trasmissione in uscita;

dendriti: le zone ricettive.

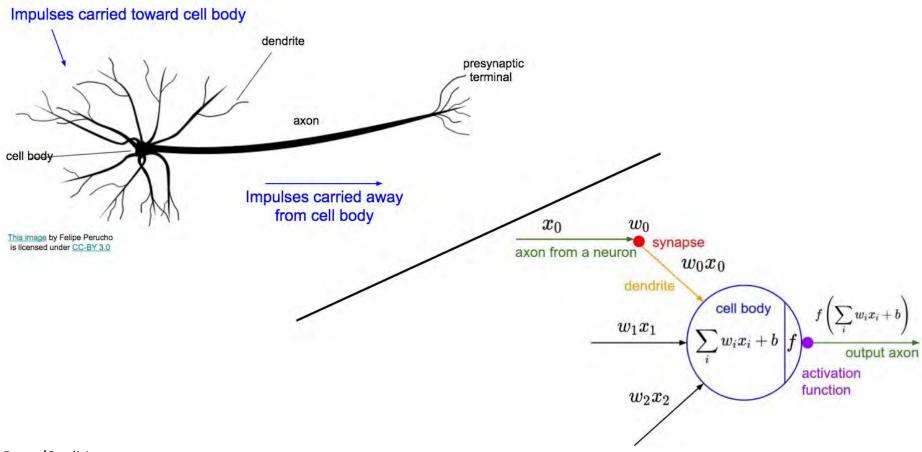
Ogni neurone riceve come input i segnali elettrici da tutti i dendriti e, se la loro composizione supera una soglia di attivazione, viene emesso un impulso elettrico in uscita verso l'assone.



Fonte/Crediti: http://www.pmar.it/pmag/ocr/ocr-08.htm



Neurone artificiale



Fonte/Crediti:

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

https://towardsdatascience.com/everything-you-need-to-know-about-activation-functions-in-deep-learning-models-84ba9f82c253



Neurone artificiale (2)

I diversi componenti sono:

 x_1 , x_2 , ..., 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 , ..., 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:

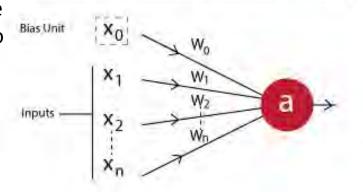
 $a = f(\sum_{i=0} w_i x_i)$

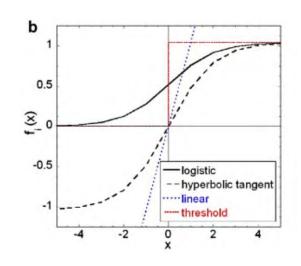
F è nota come funzione di attivazione.

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.

Fonte/Crediti: https://medium.com/hackernoon/artificial-neural-network-a843ff870338

Diagram 1: Single NN Working



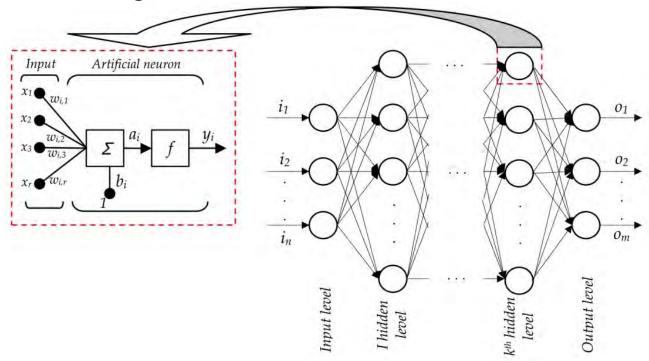




Rete neurale

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.

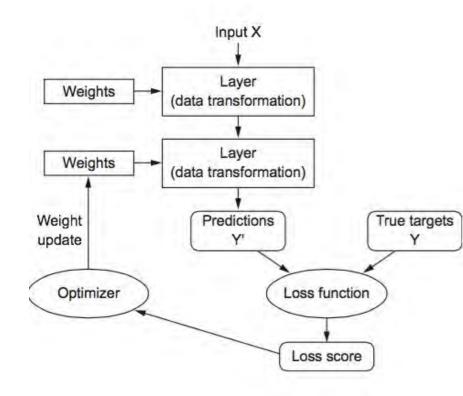


Fonte/Crediti: https://medium.com/hackernoon/artificial-neural-network-a843ff870338



Processo di apprendimento

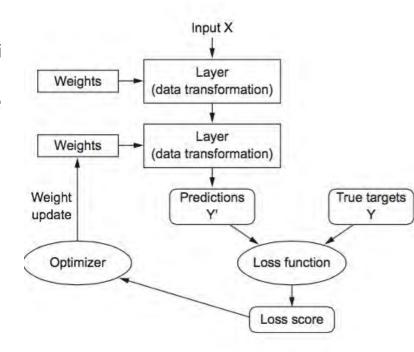
- Layers che sono combinati in una rete (o modello)
- I dati di input e gli obiettivi corrispondenti (targets)
- La funzione di costo (loss function), che definisce il segnale di retroazione utilizzato per l'apprendimento
- L'ottimizzatore (optimizer), che determina come procede l'apprendimento





Processo di apprendimento

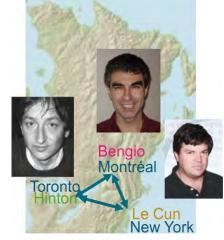
- I pesi sono fondamentali perché determinano quanto ciascun input influenzi l'output di un neurone o di un'intera rete.
- Un peso maggiore indica che l'input corrispondente è considerato più importante, mentre un peso minore indica che l'input è meno rilevante.
- I pesi consentono alla rete di apprendere 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 d apprendimento e mira a minimizzare l'errore della rete.



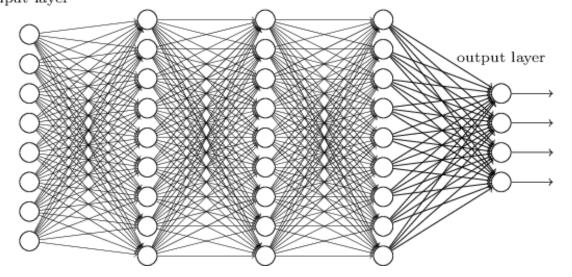


Deep Learning e reti neurali convoluzionali

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.



input layer 1 hidden layer 2 hidden layer 3



2019 per attività di ricerca su back propagation, CCN, GAN, high-dimensional word embedding and

attention, etc..

Ricercatori premiati con il

premio ACM Turing Award

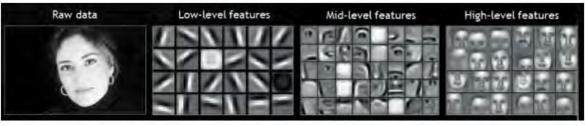
Fonte/Crediti:

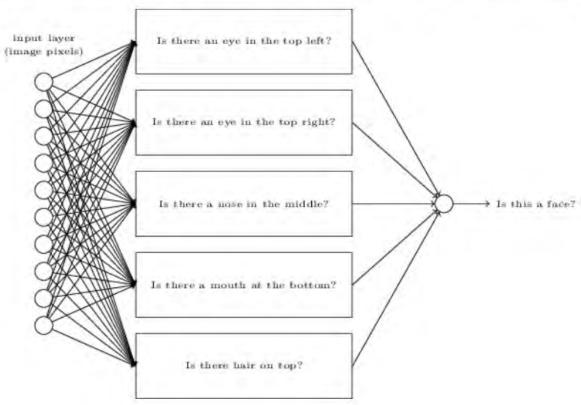
http://neuralnetworksanddeeplearning.com

https://www.slideshare.net/hammawan/deep-neural-networks



Intuizione di una rete neurali convoluzionale

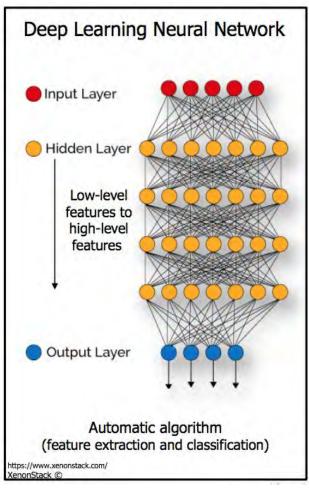


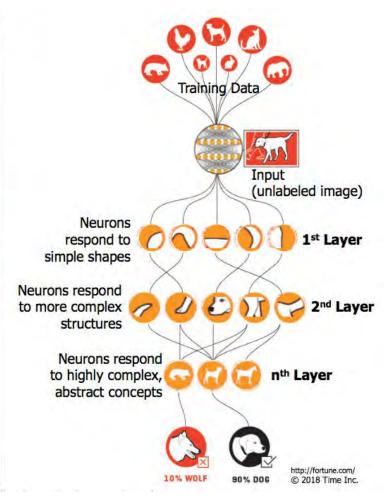


Fonte/Crediti: http://neuralnetworksanddeeplearning.com



Intuizione di una rete neurale convoluzionale





Fonte/Crediti:

https://www.xenonstack.com https://www.fortune.com



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

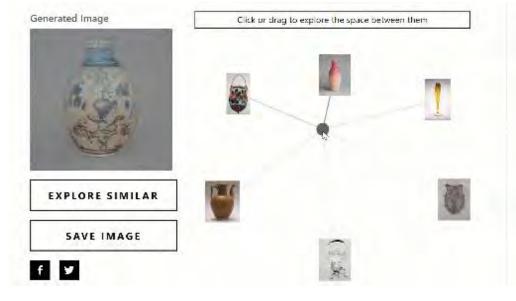


Deep learning - Arte generativa

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/Crediti:

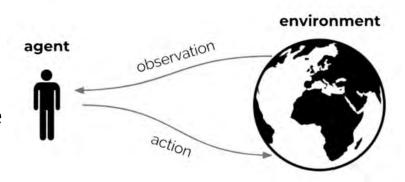
https://github.com/Microsoft/GenStudio

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
- Noi impariamo sbagliando
 (Es: imparare ad andare in bicicletta)
- Questa modalità è diversa rispetto ad altre tipologie di learning perché è:
 - Una modalità 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)



Spunto: Canale DeepMind, https://www.youtube.com/channel/UCP7jMXSY2xbc3KCAE0MHQ-A Fonte/Crediti: https://balajiai.medium.com/a-short-story-on-reinforcement-learning-e4ee1ab0ed45



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

Quando 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 raggiungere un obiettivo, 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.

Fonte/Crediti: https://axyon.ai/wp-content/uploads/2022/04/Technical-Presentation-Reinforcement-Learning-2020.pdf



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.



Fonte/Crediti: https://wallpapercrafter.com/615300-maze-graphic-render-labyrinth-design-puzzle-path-game.html

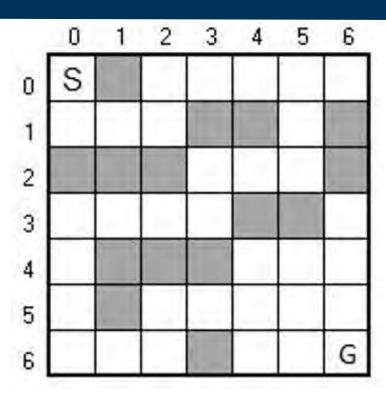


Reinforcement learning - esempio

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.

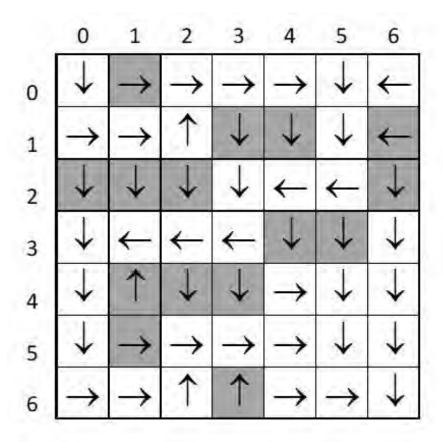


Fonte/Crediti: https://towardsdatascience.com/the-other-type-of-machine-learning-97ab81306ce9



Reinforcement learning - esempio

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



Fonte/Crediti: https://towardsdatascience.com/the-other-type-of-machine-learning-97ab81306ce9



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.

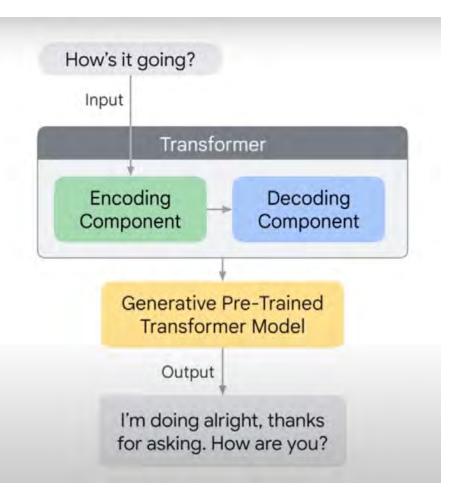


Componente non supervisionata

How it Works

Pre-Training:

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



Fonte/Crediti: **Google Cloud Tech**, Introduction to generative AI, https://www.youtube.com/watch?v=G2fqAlgmoPo



Transformers

Nel 2017 Vaswani et al. ha pubblicato un documento, "Attention is all you need", per introdurre i Transformer, una nuova classe di modelli di apprendimento automatico basati sull'attenzione.

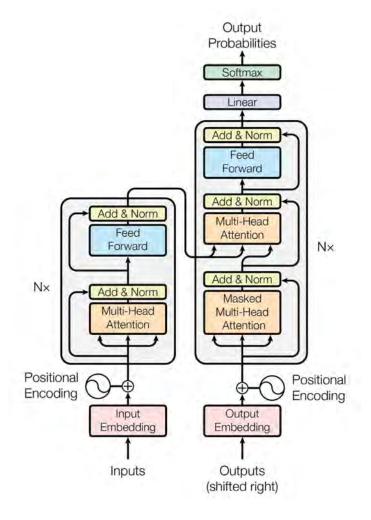
Contrariamente alle reti neurali ricorrenti, il meccanismo dell'attenzione consente di vedere l'intera frase (o anche il paragrafo) in una volta anziché in modo sequenziale una parola alla volta.

Questo permette di dare un senso alle lunghe sequenze linguistiche, creando connessioni tra diverse parti della sequenza e migliorando la comprensione del contesto in cui si trova una parola.

Fonte/Crediti:

https://en.wikipedia.org/wiki/Transformer_(machine_learning_model) Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

Cited by 88761



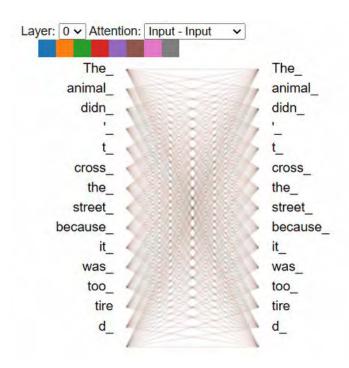


Transformers (2)

- i transformers intervengono per catturare le relazioni tra i token e l'ordine in cui compaiono nella frase.
- I pesi nella rete neurale rappresentano i parametri che determinano l'importanza relativa dei segnali di input.
 Durante l'addestramento, i pesi vengono aggiornati iterativamente in base all'errore tra l'output generato dal modello e l'output desiderato, al fine di migliorare le prestazioni della rete.

Modelli:

- GPT Generative Pre-trained Transformer
- BERT Bidirectional Encoder Representations from Transformers



Fonte/Crediti:

https://colab.research.google.com/github/tensorflow/tensor2tensor/blob/master/tensor2tensor/notebooks/hello_t2t.ipynb#scrollTo=OJKU36Q AfqOC



Addestrare un trasformer

Procedimento

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

Modelli pre-trained

L'addestramento di un modello richiede risorse importanti a livello di dataset e potenza computazionale. Si possono anche usare modelli pre addestrati (Huggingfaces, Bert di Google, i modelli GPT di OpenAI) e molto altro come i modelli preaddestrati rilasciati da Meta



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.

Gli LLM possono spesso generare output che sono falsi, tossici o semplicemente non utili per l'utente.

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

Open Al nel 2017 ha sviluppato un algoritmo di apprendimento per rinforzo dal feedback umano 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). Il feedback umano viene raccolto chiedendo agli umani di classificare le istanze del comportamento dell'agente.

Fonte/Crediti:

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

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



Supervised fine tuning - Reinforcement learning from human feedback

Step 1

Collect demonstration data, and train a supervised policy.

A prompt is sampled from our prompt dataset.

A labeler demonstrates the desired output behavior.

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



Step 2

Collect comparison data, and train a reward model.

A prompt and several model outputs are sampled.

A labeler ranks the outputs from best to worst.

This data is used to train our reward model.









Step 3

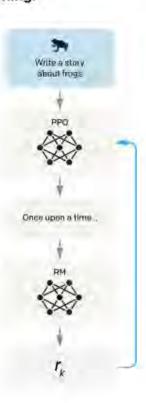
Optimize a policy against the reward model using reinforcement learning.

A new prompt is sampled from the dataset.

The policy generates an output.

The reward model calculates a reward for the output,

The reward is used to update the policy using PPO.



Fonte/Crediti: https://openai.com/blog/chatgpt



Cosa c'è dietro l'angolo

Autonomous Al Generative agents



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.



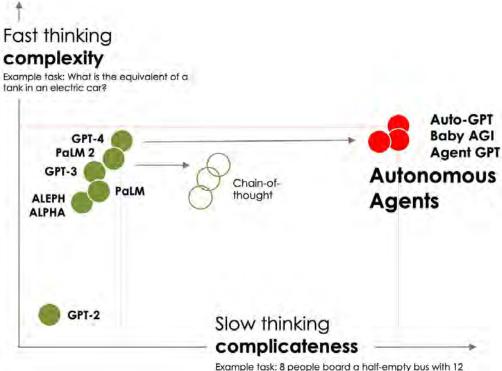
Autonomous agents (2)

Gli agenti autonomi (autonomous agents) sono invece modelli che 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



Example task: 8 people board a half-empty bus with 12 rows of 4 seats each. At the next stop, 17 get off and 9 get on. Then half of the passengers get off and 3 get on. How many seats are empty when all passengers are seated?

Fonte/Crediti:

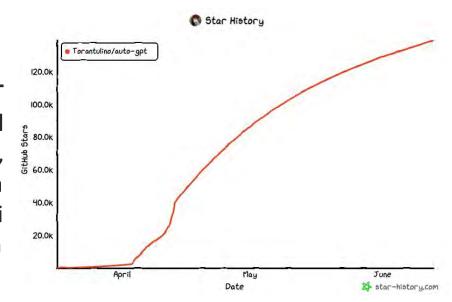
Shen, Y., Song, K., Tan, X., Li, D., Lu, W., & Zhuang, Y. (2023). **Hugginggpt: Solving ai tasks with chatgpt and its friends in huggingface.** *arXiv preprint arXiv:2303.17580*.



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 interazione 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 <u>istanze</u> <u>ricorsive</u> e in risposta alle nuove informazioni acquisite.



Fonte/Crediti: 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)

Fonte/Crediti: screenshot da interrogazione servizio https://agentgpt.reworkd.ai/



AgentGPT (2)

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

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

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

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

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

Fonte/Crediti: screenshot da interrogazione servizio https://agentgpt.reworkd.ai/



HuggingGPT

25 Maggio 2023

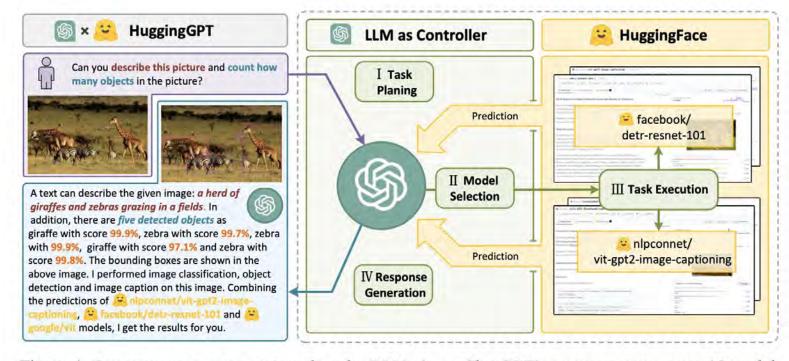


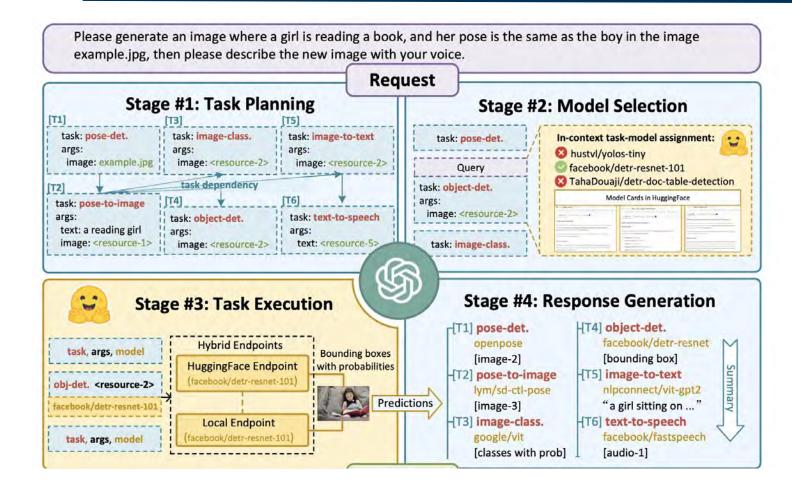
Figure 1: Language serves as an interface for LLMs (e.g., ChatGPT) to connect numerous AI models (e.g., those in Hugging Face) for solving complicated AI tasks. In this concept, an LLM acts as a controller, managing and organizing the cooperation of expert models. The LLM first plans a list of tasks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

Fonte/Crediti: 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*



HuggingGPT

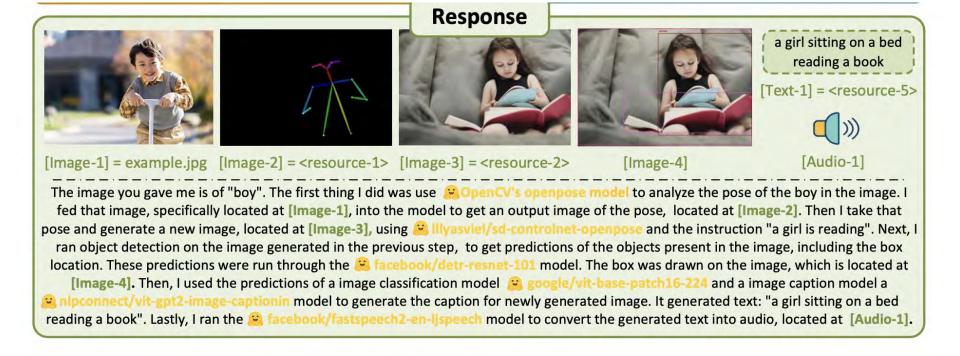
25 Maggio 2023



Fonte/Crediti: 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*



HuggingGPT



Fonte/Crediti: 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*



"Westworld lured millions of us into a fantasy of human-like robots in a Wild West-themed world. Could it be real one day?

Stanford/Google researchers just told us how they used AI to make "generative agents."



Figure 1: Generative agents create believable simulacra of human behavior for interactive applications. In this work, we demonstrate generative agents by populating a sandbox environment, reminiscent of The Sims, with twenty-five agents. Users can observe and intervene as agents they plan their days, share news, form relationships, and coordinate group activities.

Fonte/Crediti: 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.

https://www.youtube.com/watch?v=HPHwN5ZxF98 25 mag 2023

https://reverie.herokuapp.com/arXiv_Demo/





Fonte/Crediti: 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.

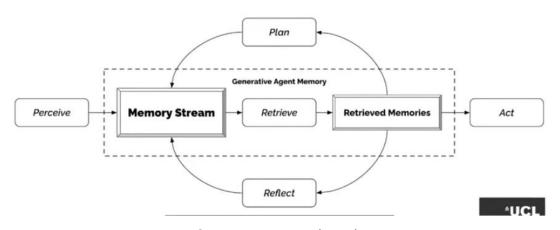
https://www.youtube.com/watch?v=HPHwN5ZxF98 25 mag 2023

https://reverie.herokuapp.com/arXiv_Demo/



How can we leverage the power of generative models to simulate believable human behavior?

A novel agent architecture that remembers, reflects, and plans based on constantly growing memories and cascading social dynamics.



Fonte/Crediti: 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.

https://www.youtube.com/watch?v=HPHwN5ZxF98 25 mag 2023

https://reverie.herokuapp.com/arXiv Demo/



- Gli agenti generativi si svegliano, cucinano la colazione e si dirigono al lavoro; gli artisti dipingono, mentre gli scrittori scrivono; formano opinioni, si notano reciprocamente e iniziano conversazioni; ricordano e riflettono sui giorni passati mentre pianificano il giorno successivo.
- Unendo grandi modelli linguistici con agenti computazionali interattivi, questo paper introduce schemi architettonici e di interazione per abilitare simulazioni credibili del comportamento umano.
- Ad esempio, partendo solo da una singola idea specificata dall'utente, che un agente desidera organizzare una festa di San Valentino, gli agenti diffondono autonomamente inviti alla festa nei due giorni successivi, fanno nuove conoscenze, si chiedono reciprocamente di uscire per la festa e si coordinano per presentarsi insieme alla festa al momento giusto.

Fonte/Crediti: 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.

https://www.youtube.com/watch?v=HPHwN5ZxF98 25 mag 2023

https://reverie.herokuapp.com/arXiv Demo/



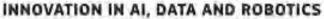
Normativa e attività della UE



Al nei progetti europei

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

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



Al, Data and Robotics for the Green Deal

INNOVATION IN AI, DATA AND ROBOTICS

Al, Data and Robotics at work

LEADERSHIP IN AI BASED ON TRUST

Tackling gender, race and other biases in Al

LEADERSHIP IN AI BASED ON TRUST

Verifiable robustness, energy efficiency and transparency for Trustworthy Al:

Scientific excellence boosting industrial competitiveness

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



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



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/Crediti: 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/Crediti: 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



Scenari e rischi



MOLTO PROBABILMENTE

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

POCO PROBABILMENTE

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

Fonte: Al takes center stage, Bot.Me: A revolutionary partnership, Pwc 2017

QUALI POSSONO ESSERE I POTENZIALI DANNI ASSOCIATI A SISTEMI IA

FISICI POLITICI DIGITALI Phishing automatizzato, Terrorismo automatizzato che Propaganda, attraverso creazione di email false, siti web utilizza droni o veicoli autonomi immagini e video falsi e link per sottrarre informazioni. come armi. facilmente generabili. Hacking massivo, attraverso la Sciami di Robot che cercano di Rimozione automatica del scoperta automatizzata delle raggiungere lo stesso obiettivo. dissenso, grazie alla possibilità vulnerabilità dei sistemi. di trovare, analizzare e rimuovere automaticamente Attacchi remoti, resi possibili Ingannare sistemi IA sfruttando dal fatto che i robot autonomi testi ed immagini. a proprio vantaggio i difetti possono essere controllati da nell'interpretazione della realtà qualunque distanza. Persuasione personalizzata, da parte dell'IA. con l'utilizzo di informazioni pubbliche per influenzare l'opinione di qualcuno. Fonte: Al experts list the real dangers of artificial intelligence, Dave Gershgorn Feb. 2018

Fonte/Crediti: 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.



Digital Service Act e Al

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 per migliorare la trasparenza dei sistemi di raccomandazione e della pubblicità.

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

Fonte/Crediti: https://digital-strategy.ec.europa.eu/it/policies/digital-services-act-package



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.

Fonte/Crediti: **ECAT**, ttps://algorithmic-transparency.ec.europa.eu/index_en



AI ACT

Il Parlamento europeo in data 14 giugno 2023 ha dato il via libera all'**Artificial Intelligence Act**, che regolerà l'Intelligenza Artificiale nel rispetto dei diritti e dei valori dell'Unione Europea. 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'Al progettata per essere neutrale rispetto alla tecnologia, in modo che possa applicarsi ai sistemi di Al di oggi e di domani.

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

Fonte/Crediti:

https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_IT.html

https://www.ilsole24ore.com/art/il-parlamento-europeo-approva-l-ai-act-cosa-cambiera-le-nostre-aziende-AE2aw3gD



Al ACT - misure di trasparenza

I membri del Parlamento Europeo hanno incluso obblighi per i fornitori di modelli di Al 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.

HOME / TECNOLOGIA

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

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

Fonte/Crediti:

https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_IT.html

https://www.ilsole24ore.com/art/il-parlamento-europeo-approva-l-ai-act-cosa-cambiera-le-nostre-aziende-AE2aw3gD



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

La nuova legge promuove le aree di sperimentazione regolamentate, o ambienti controllati, istituiti dalle autorità pubbliche per testare l'Al 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.

Fonte/Crediti:

https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_IT.html

https://www.ilsole24ore.com/art/il-parlamento-europeo-approva-l-ai-act-cosa-cambiera-le-nostre-aziende-AE2aw3gD



Strategia di ENISA per la Al

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/Crediti: ENISA AI CYBERSECURITY CONFERENCE - 7 giugno 2023

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



Segnali di allarme

Geoffrey Hinton uno dei Nobel per l'informatica del 2018 ha deciso nel mese di aprile 2023 di lasciare Google per poter intervenire sui rischi dell'IA senza danneggiare la reputazione di Google

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.



Fonte/Crediti:

Il Manifesto, https://ilmanifesto.it/lidea-piu-pericolosa-dei-padri-di-ai-e-che-non-ci-sia-via-di-scampo Financial Times, https://www.ft.com/content/c2b0c6c5-fe8a-41f2-a4df-fddba9e4cd88



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 OpenAl, 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: Other Notable Figures Al Scientists **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, OpenAl Dario Amodei CEO. Anthropic Dawn Song Professor of Computer Science, UC Berkeley Ya-Qin Zhang Professor and Dean, AIR, Tsinghua University Ilva Sutskever Co-Founder and Chief Scientist, OpenAl 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

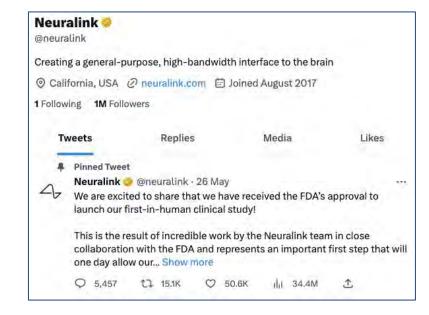
Fonte/Crediti: Safe AI, https://safe.ai/statement-on-ai-risk



La ricerca non si arresta

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







Conclusione

"La corrente ondata di intelligenza artificiale generativa è una sotto-categoria 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/Crediti: https://hai.stanford.edu/sites/default/files/2023-03/Generative_AI_HAI_Perspectives.pdf marzo 2023



Grazie per l'attenzione. Domande?

