

L'intelligenza artificiale: definizioni etica strategie prospettive



***Roberto Magnani
Milano 4 Luglio 2023***

Intelligenza artificiale

I padri



Alan Turing
(1912-1954)

formalizzazione dei
concetti di algoritmo e
calcolo mediante
l'omonima macchina



Claude Shannon,
(1916 -2001)
il padre della teoria
dell'informazione,



Marvin Minsky,
(1927 -2016)
matematico informatico
fondatore del laboratorio di
IA al MIT di Boston,

Dartmouth College 1956,



John McCarthy
tra fondatori e
contributori allo
sviluppo della
robotica

Classificazione storica Intelligenza Artificiale

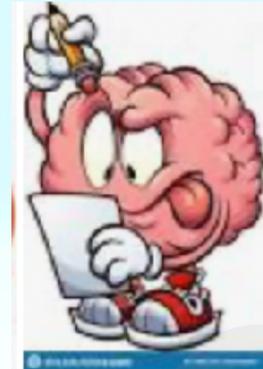
● RISTRETTA



Riconoscimento immagini
Vettura a guida indipendente

ANI - Artificial Narrow Intelligence
basata su algoritmi specificatamente
dedicati a un compito

● GENERALE



GPT4... forse ???

AGI - Artificial General Intelligence che
cerca di riprodurre i comportamenti della
mente umana

• SUPERINTELLIGENZA



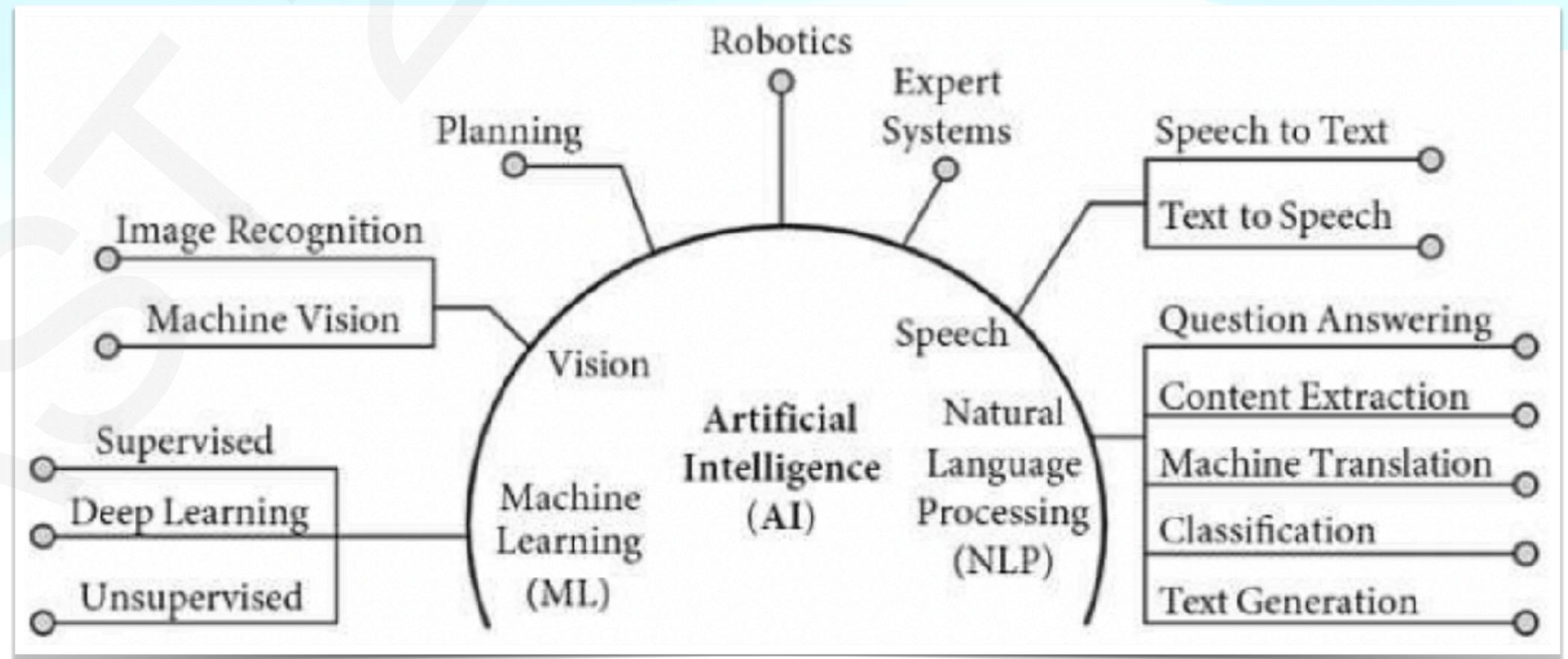
ASI - Artificial SuperIntelligence in grado
di avere a disposizione, capire, imparare ed
utilizzare l'intero scibile umano

HAL - 2001 Odissea nello spazio

Intelligenza artificiale

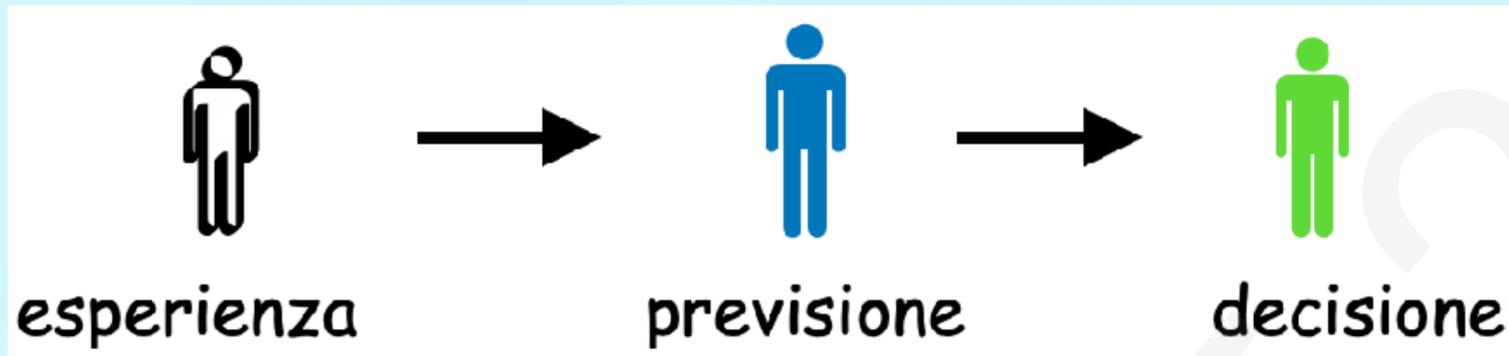
Definizioni e aree di lavoro

abilità di una macchina di mostrare capacità umane quali il ragionamento, l'apprendimento, la pianificazione e la creatività.

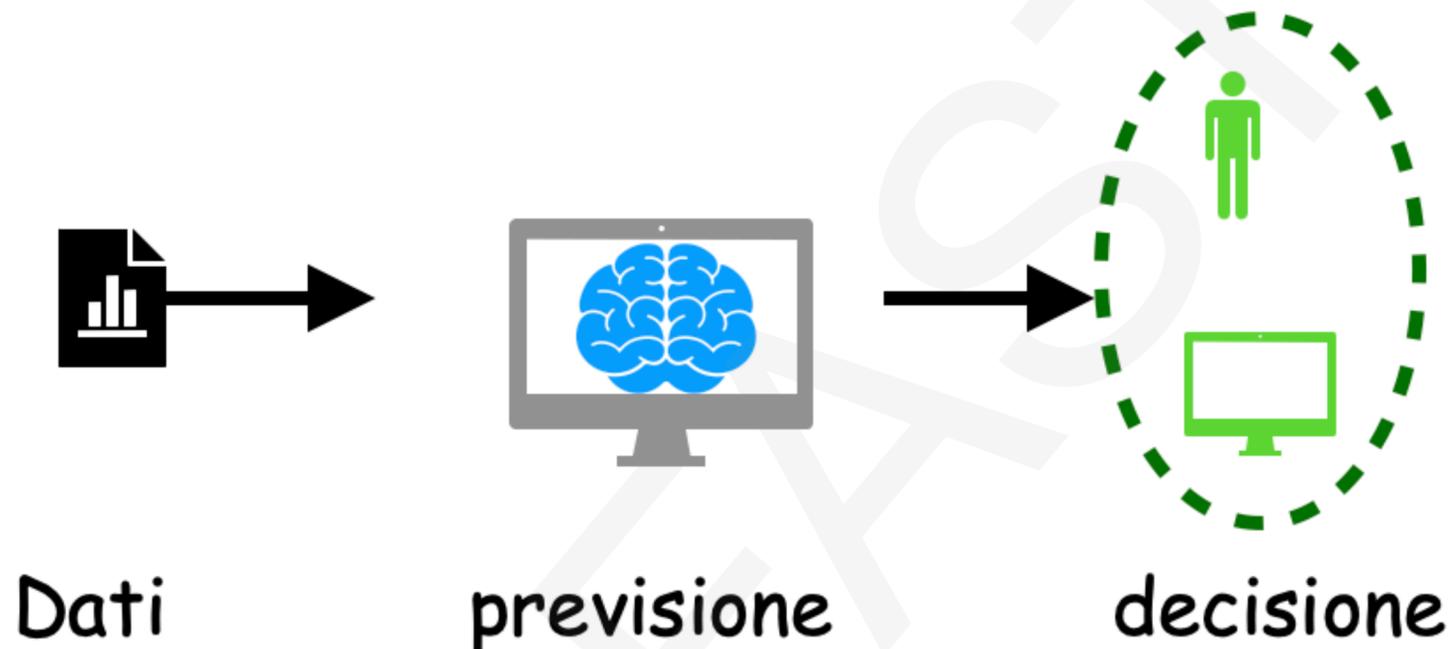


Apprendimento automatico

esperienza e dati



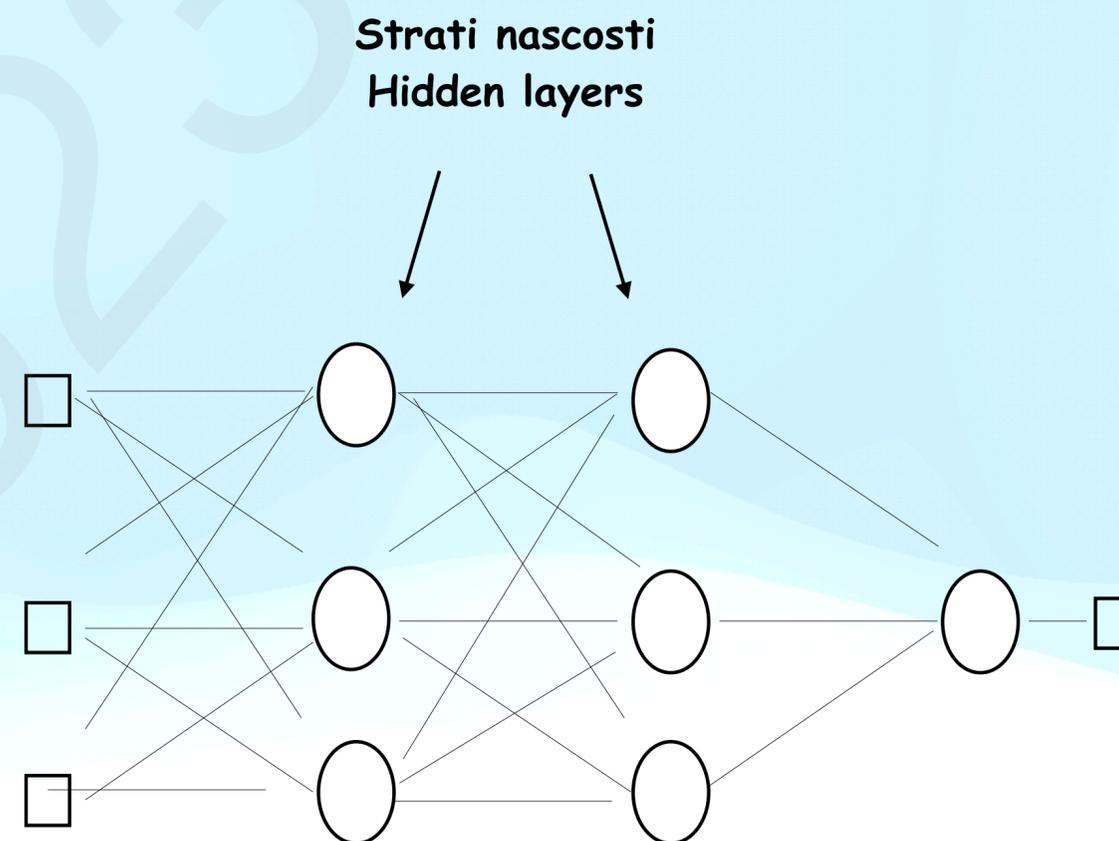
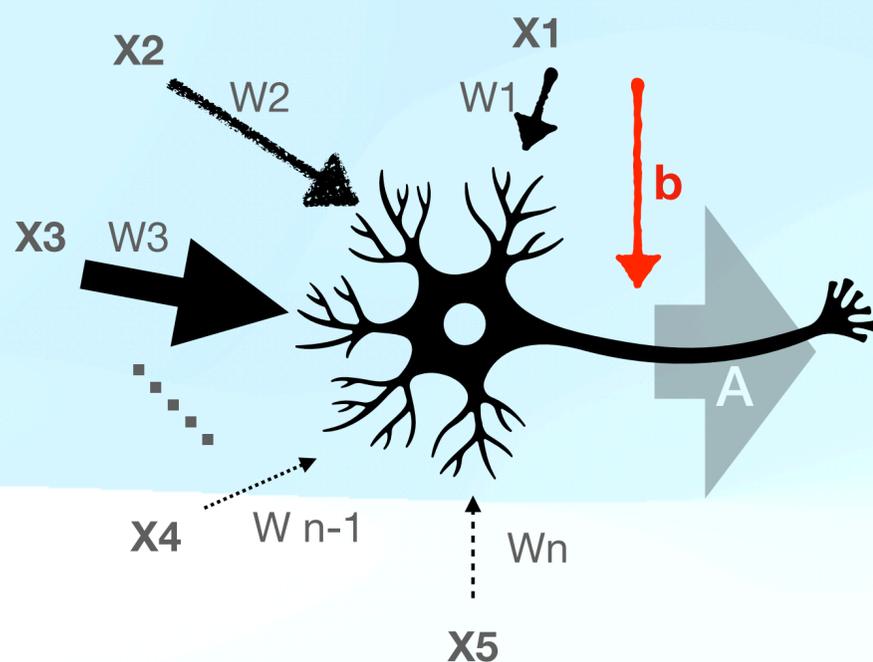
L'apprendimento automatico è semplicemente l'imparare dai dati e il trasformarli in previsioni.



Proprio come impariamo dall'esperienza, la macchina impara dai dati

Le reti di neuronali

I singoli elementi effettuano somme



I pesi (w) e "bias" (b) sono i parametri, essendo regolabili, sono il mezzo attraverso il quale una rete di neuroni impara.

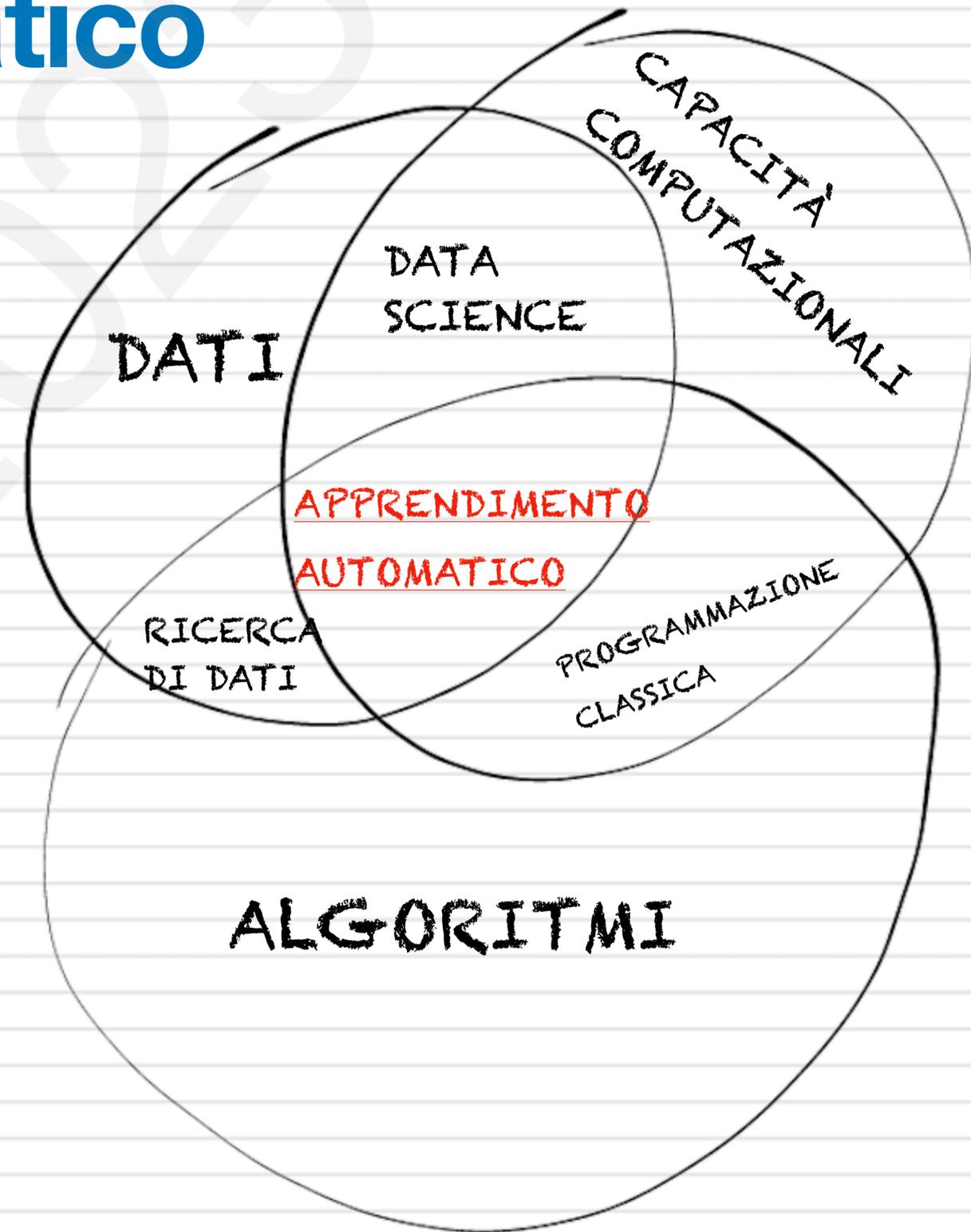
La funzione di attivazione (A) conferisce alla rete di neuroni la capacità di esprimersi, in genere in modo non lineare.

Si parla di "deep-learning" o rete neurale quando i livelli di neuroni nascosti sono ≥ 3

Apprendimento automatico

Dati + Algoritmi + Calcolo

- **Crescita esponenziale dei dati disponibili**
 - 90% dei dati mondiali è stato generato negli ultimi due anni.
 - entro il 2025, il 50% dei dati generati e trasmessi saranno di tipo M2M
- **Sviluppo Algoritmi.**
 - Apprendimento automatico, visione digitale, robotica, elaborazione del linguaggio, pianificazione e l'apprendimento basati su agenti, la sicurezza dell'IA, la privacy e la trasparenza dell'IA
- **Capacità di calcolo**
 - Supercomputer / Computer quantistico



Tipi di apprendimento automatico

- Apprendimento Supervisionato



- Apprendimento non Supervisionato

- Apprendimento per rinforzo



- creazione del modello con supervisione umana

- *Si attribuiscono “etichette” ai dati usati per l’addestramento del modello*

- *La macchina impara dagli esempi ed elabora un modello predittivo*

- non è guidato in fase di addestramento

- *utile quando è necessario identificare modelli e utilizzare i dati per **prendere decisioni**.*

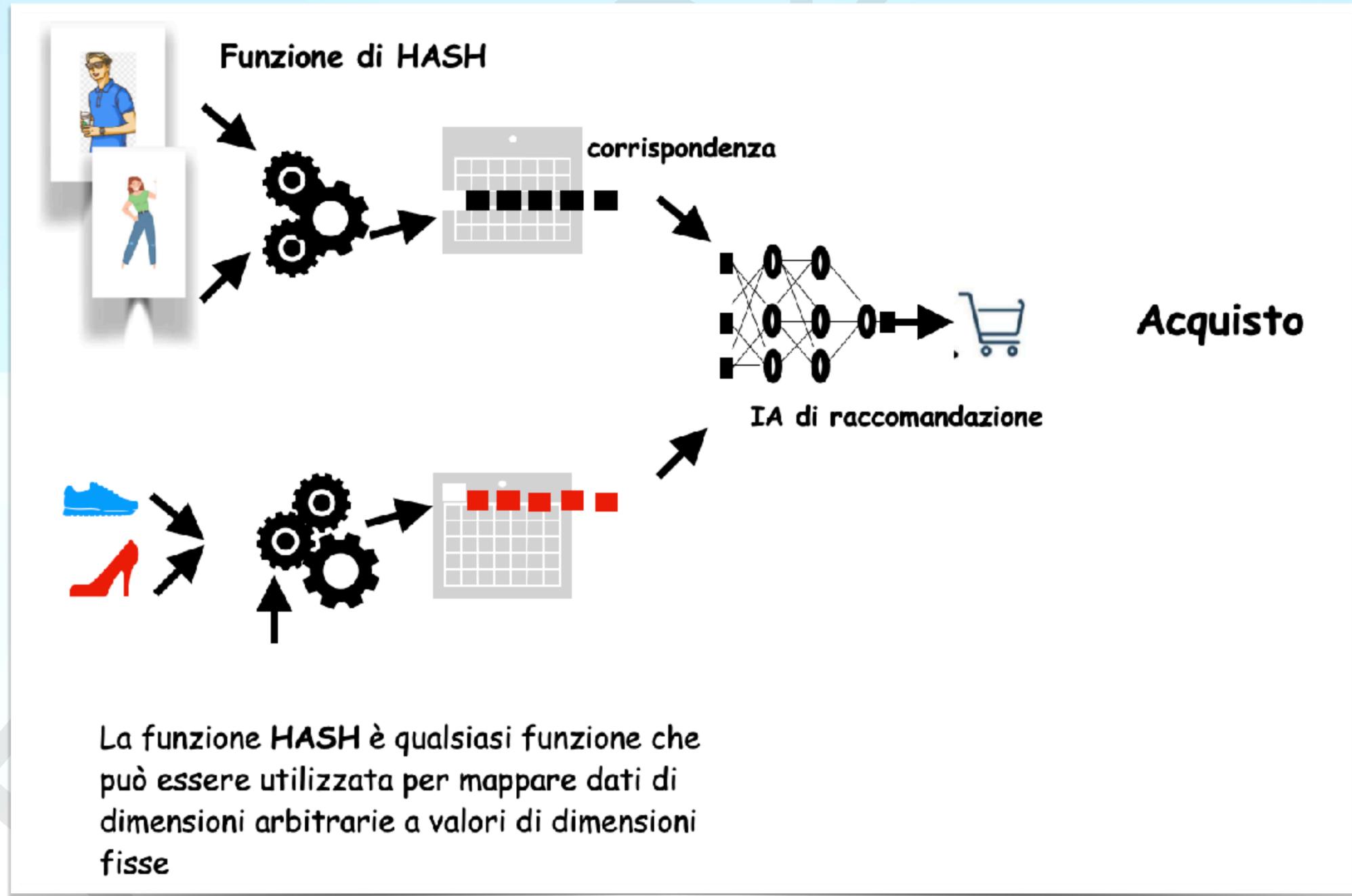
- Agenti autonomi in grado di decidere azioni da eseguire per il raggiungimento di obiettivi prefissati

- decisioni sequenziali, in cui l'azione da compiere dipende dallo stato attuale del sistema e ne determina quello futuro

- La qualità di un'azione è data da un valore numerico di “**ricompensa**” o valutazione

Esempio semplificato

Schema di massima proposta di prodotto



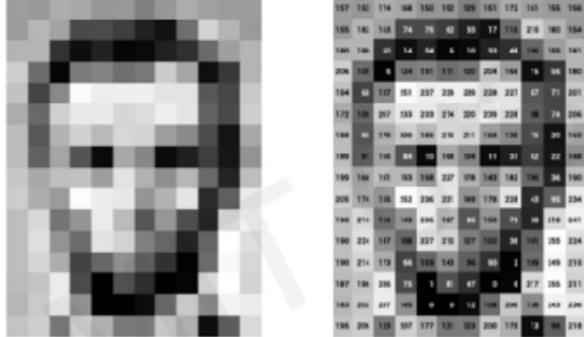
Le immagini sono numeri

Con regressione definisco valori continui con classificazione attribuisco probabilità

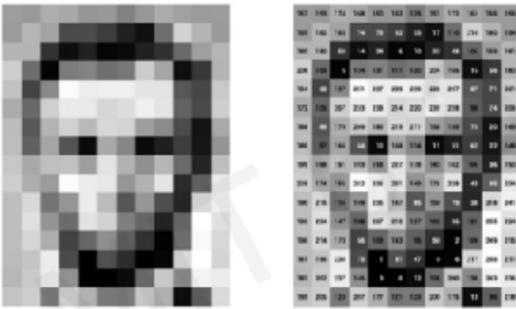
Images are Numbers



Images are Numbers



Images are Numbers



What the computer sees

107	145	174	144	140	142	120	97	77	62	56	54
155	145	110	74	75	62	39	17	110	210	180	154
140	148	81	14	44	4	54	64	44	140	144	140
208	151	8	224	121	171	220	204	184	74	14	140
184	54	127	151	237	239	209	238	227	27	71	201
173	130	207	133	233	274	200	230	228	16	74	206
144	44	176	206	140	176	211	144	134	16	14	144
188	51	149	84	140	138	134	11	21	12	22	148
188	144	111	163	134	237	178	143	182	140	36	180
205	174	116	163	204	221	140	170	224	43	162	224
144	214	174	148	844	147	81	144	71	16	174	441
180	224	117	166	227	210	127	122	34	111	258	224
140	214	173	64	133	143	74	16	1	126	249	214
187	136	236	74	1	47	0	1	27	254	211	
144	214	417	144	1	14	144	204	144	414	244	
140	204	132	107	177	121	124	204	174	13	16	214

An image is just a matrix of numbers $[0,255]$
i.e., $1080 \times 1080 \times 3$ for an RGB image

Tasks in Computer Vision



Input: Image

Pixel Representation

classification

Lincoln	0.8
Washington	0.1
Jefferson	0.05
Obama	0.05

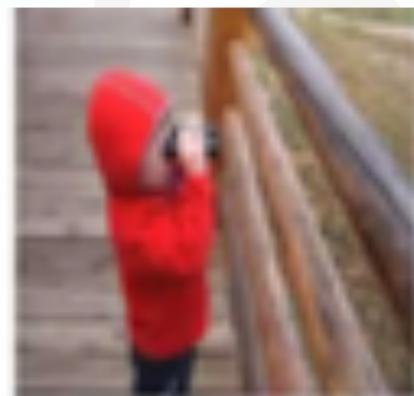
- **Regression:** output variable takes continuous value
- **Classification:** output variable takes class label. Can produce probability of belonging to a particular class

Il riconoscimento facciale avrà
 forti limitazioni nella Comunità
 Europea secondo AI ACT del 14
 Giugno 2023

Criticità per la visione computerizzata

- creare un **dataset** sufficientemente ampio per l'addestramento dell'algoritmo;
- insegnare all'algoritmo a comprendere l'immagine anche in presenza di **trasformazioni** (es. condizioni di luminosità non ottimale, deformazione o copertura parziale del soggetto, variazioni di scala).

Bimbo con binocolo
Sotto la pioggia



Filtraggio di pixel
in archivio



Bimbo confuso con
zoccolo rosso

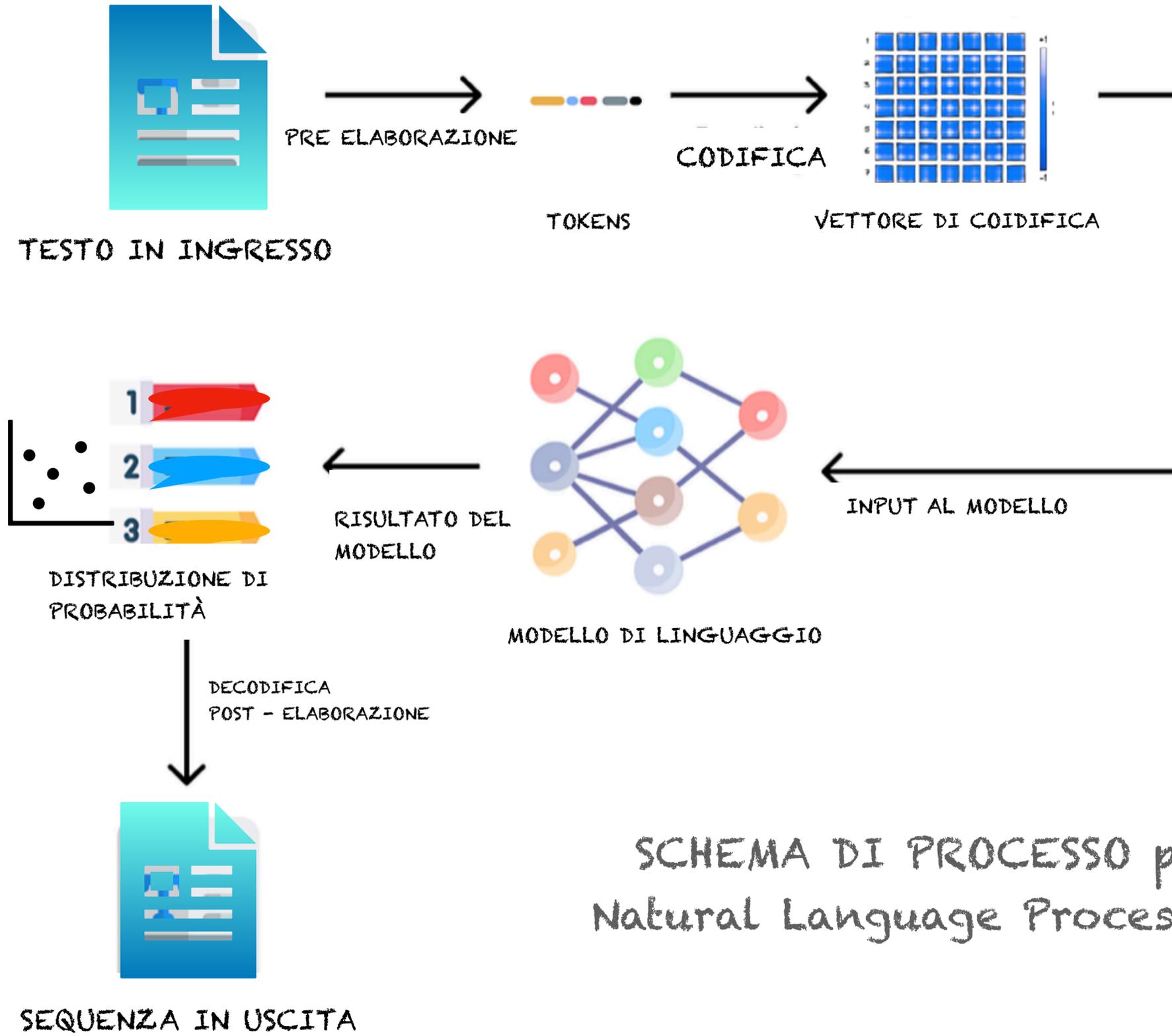


NLP

Da un Corpus costituito da dati disponibili si procede con classificazioni e processi statistici



Esempio -GPT 3 base di addestramento



SCHEMA DI PROCESSO per
Natural Language Processing

Il cuore delle IA generative - il “trasformer”

Sviluppato da Google e Università di Toronto per favorire le traduzioni

I punti salienti

Codifica posizionale
(Positional encoding)

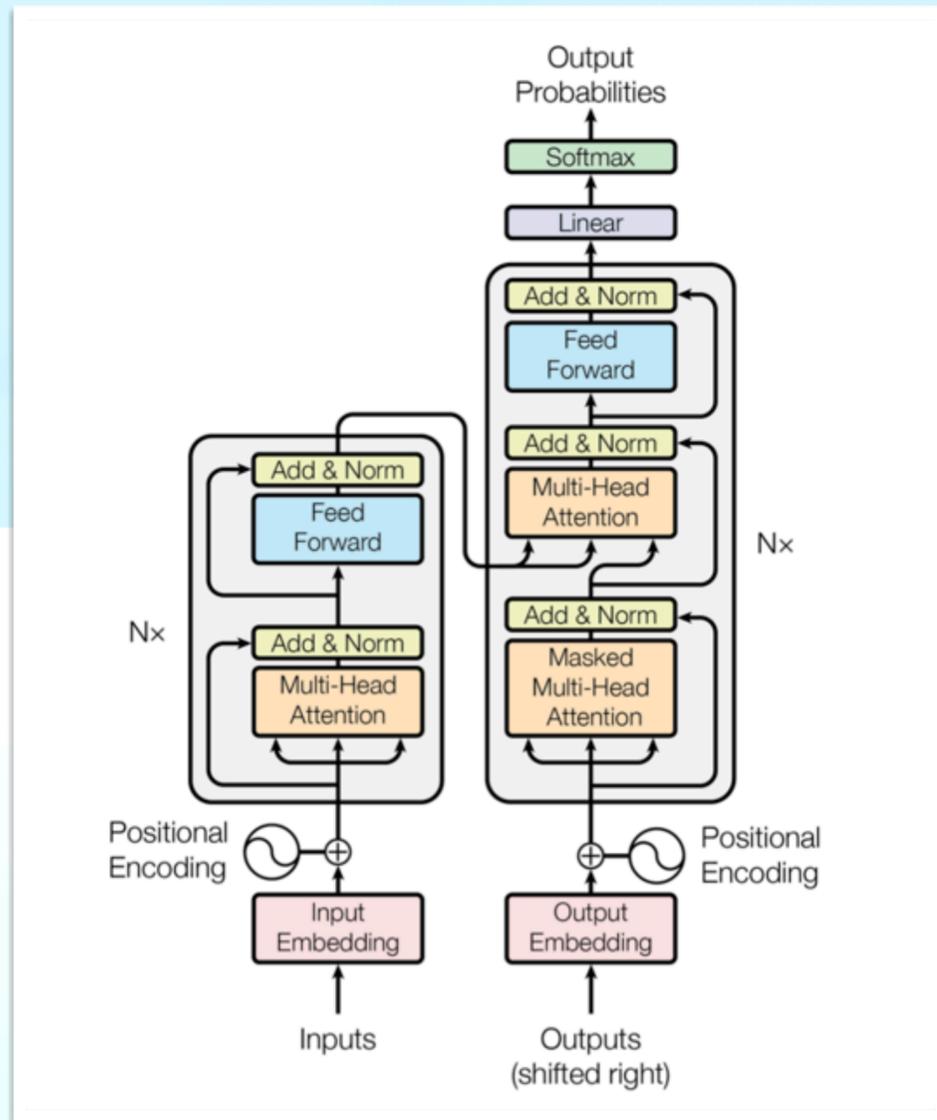
Attenzione

Auto-Attenzione
(Self Attention)

Positional encoding fornisce l'informazione della **posizione nella struttura dei dati** invece che nella rete. Il sistema impara l'importanza dell'ordine direttamente dai dati. Facilita l'addestramento del sistema

“attenzione” da il titolo all'articolo. Fondamentale nelle traduzioni, individua la **relazione tra parole lontane in una frase**, o valuta l'importanza ad gruppi di parole proveniente da testi diversi (cross attention)

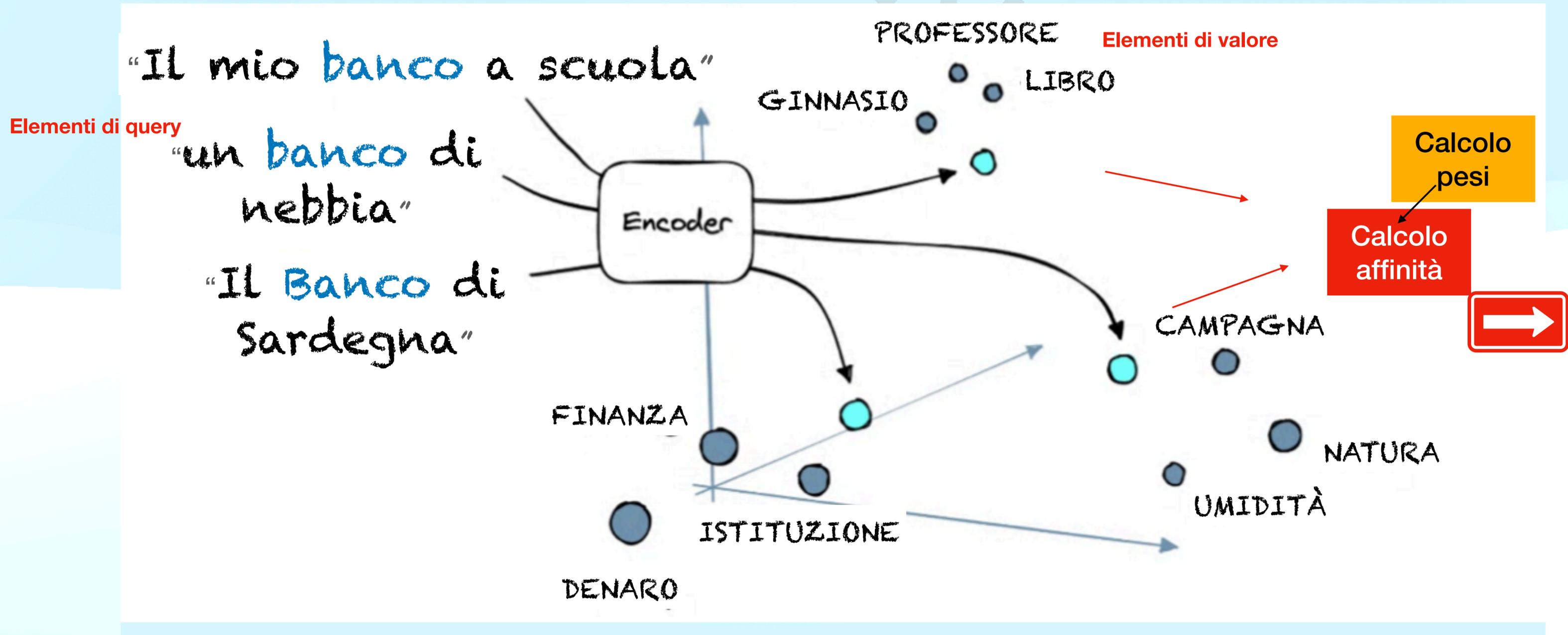
L'auto-attenzione è la novità del trasformatore, **calcola l'importanza delle singole parole**. Attribuisce un punteggio di attenzione per le altre parole nella frase. Aiuta a calcolare le rappresentazioni di ogni parola (token) in base alle altre nella stessa frase. Dà significati diversi a frasi con le stesse parole ma con posizioni diverse.



Dall'articolo "Attention is all you need"
Ashish Vaswani e at. Dicembre 2017

Cross Attention

Esempio



2 Scope and Limitations of this Technical Report

This report focuses on the capabilities, limitations, and safety properties of GPT-4. GPT-4 is a Transformer-style model [39] pre-trained to predict the next token in a document, using both publicly available data (such as internet data) and data licensed from third-party providers. The model was then fine-tuned using Reinforcement Learning from Human Feedback (RLHF) [40]. Given both the competitive landscape and the safety implications of large-scale models like GPT-4, this report contains no further details about the architecture (including model size), hardware, training compute, dataset construction, training method, or similar.

We are committed to independent auditing of our technologies, and shared some initial steps and ideas in this area in the system card accompanying this release.² We plan to make further technical details available to additional third parties who can advise us on how to weigh the competitive and safety considerations above against the scientific value of further transparency.

3 Predictable Scaling

A large focus of the GPT-4 project was building a deep learning stack that scales predictably. The primary reason is that for very large training runs like GPT-4, it is not feasible to do extensive model-specific tuning. To address this, we developed infrastructure and optimization methods that have very predictable behavior across multiple scales. These improvements allowed us to reliably predict some aspects of the performance of GPT-4 from smaller models trained using $1,000\times$ – $10,000\times$ less compute.

Esempi di uso di modelli generativi con immagini

Guida automatica

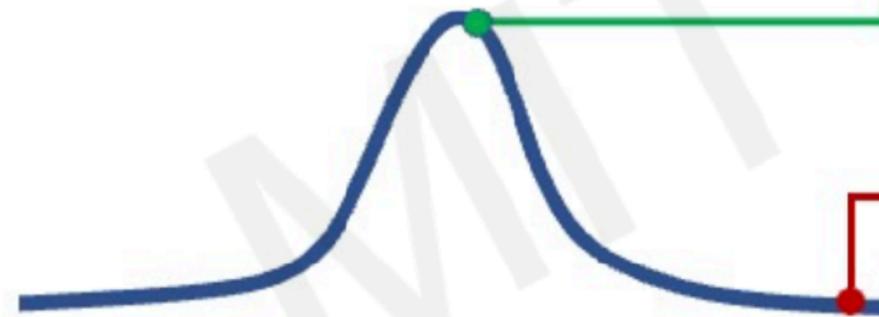
Why generative models? Outlier detection

- **Problem:** How can we detect when we encounter something new or rare?
- **Strategy:** Leverage generative models, detect outliers in the distribution
- Use outliers during training to improve even more!

95% of Driving Data:
(1) sunny, (2) highway, (3) straight road



Detect outliers to avoid unpredictable behavior when training



Edge Cases



Harsh Weather

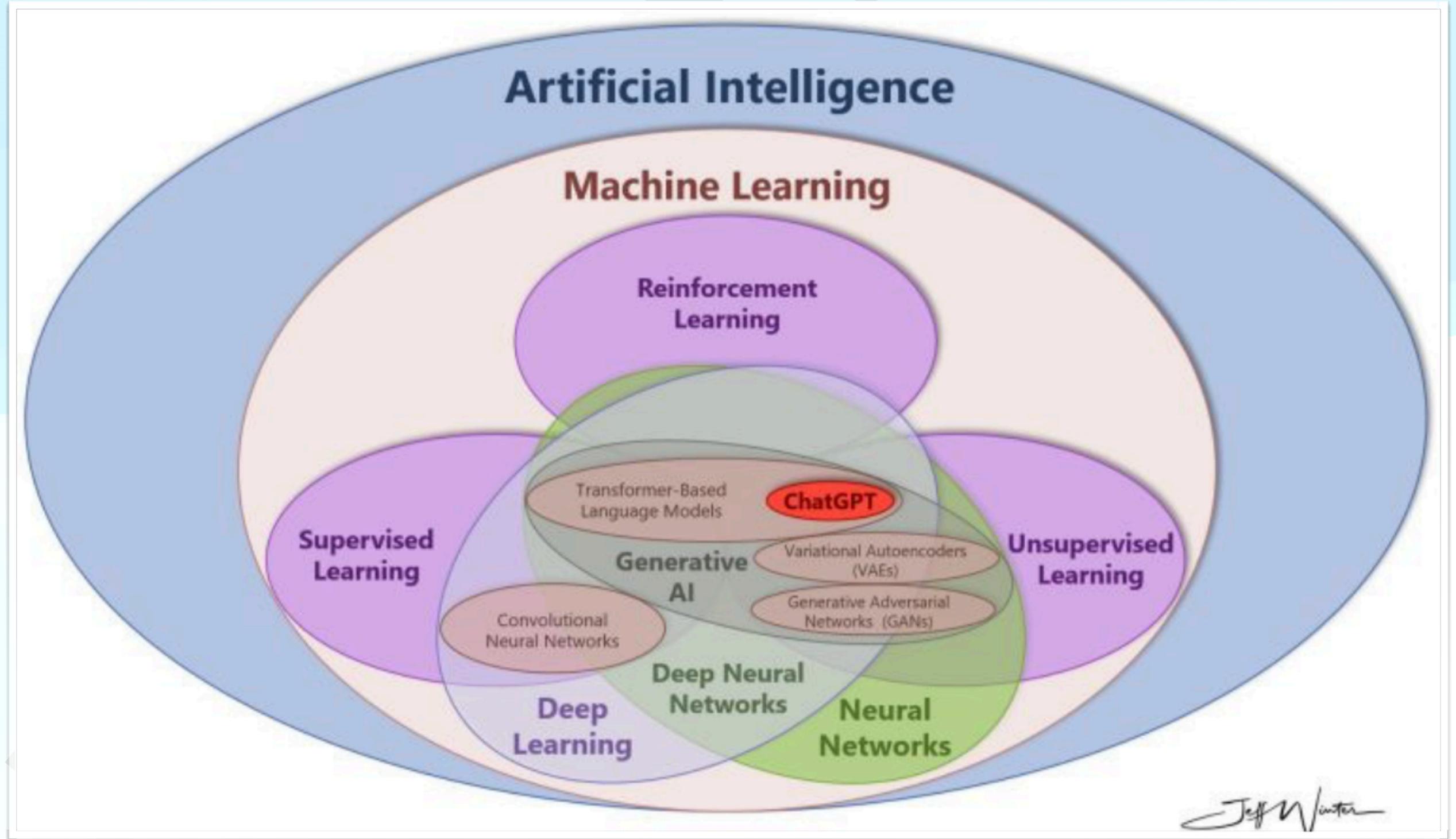


Pedestrians

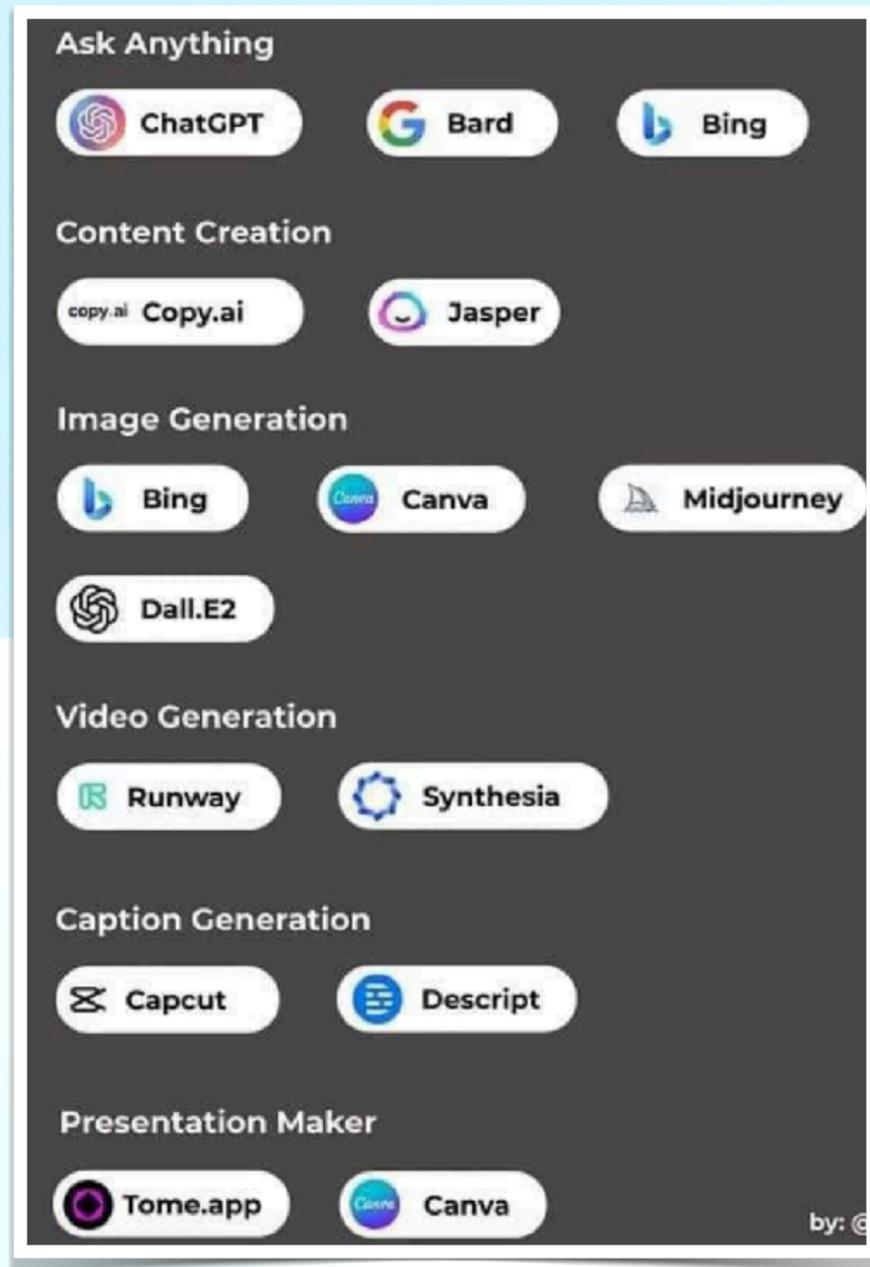
courtesy MIT 'Deep learning introduction'

Le tecniche in uso

Sommario



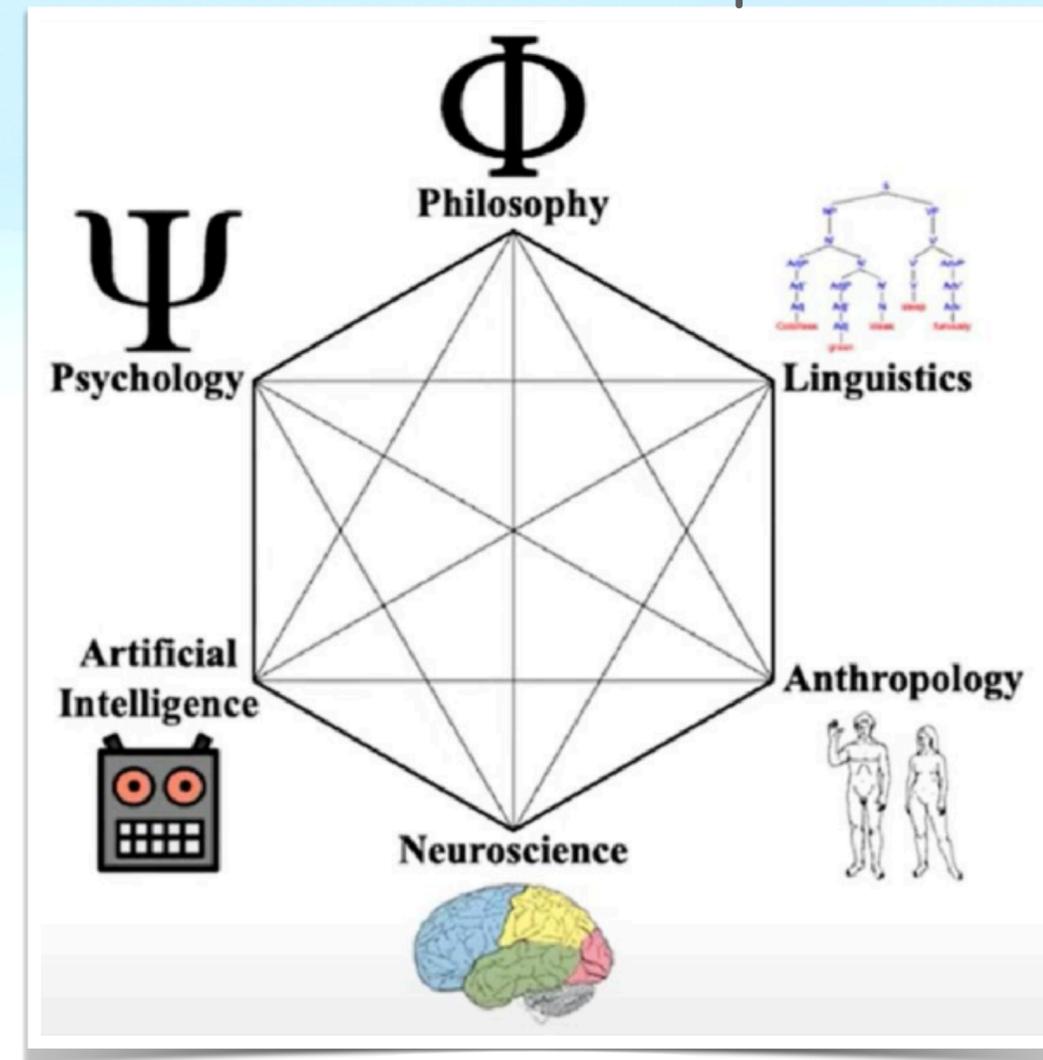
L'utilizzo di IA generative nei vari campi



Lista indicativa non esaustiva ne valutazione di performance



Non solo tecnologia ma un supporto a molte altre discipline



Il problema della sostenibilità delle IA generative

1a generazione : Trasformers BERT, DistillBERT,

Attention, parallelizzabili, architettura encoder-decoder

1 miliardo di parametri

2a generazione

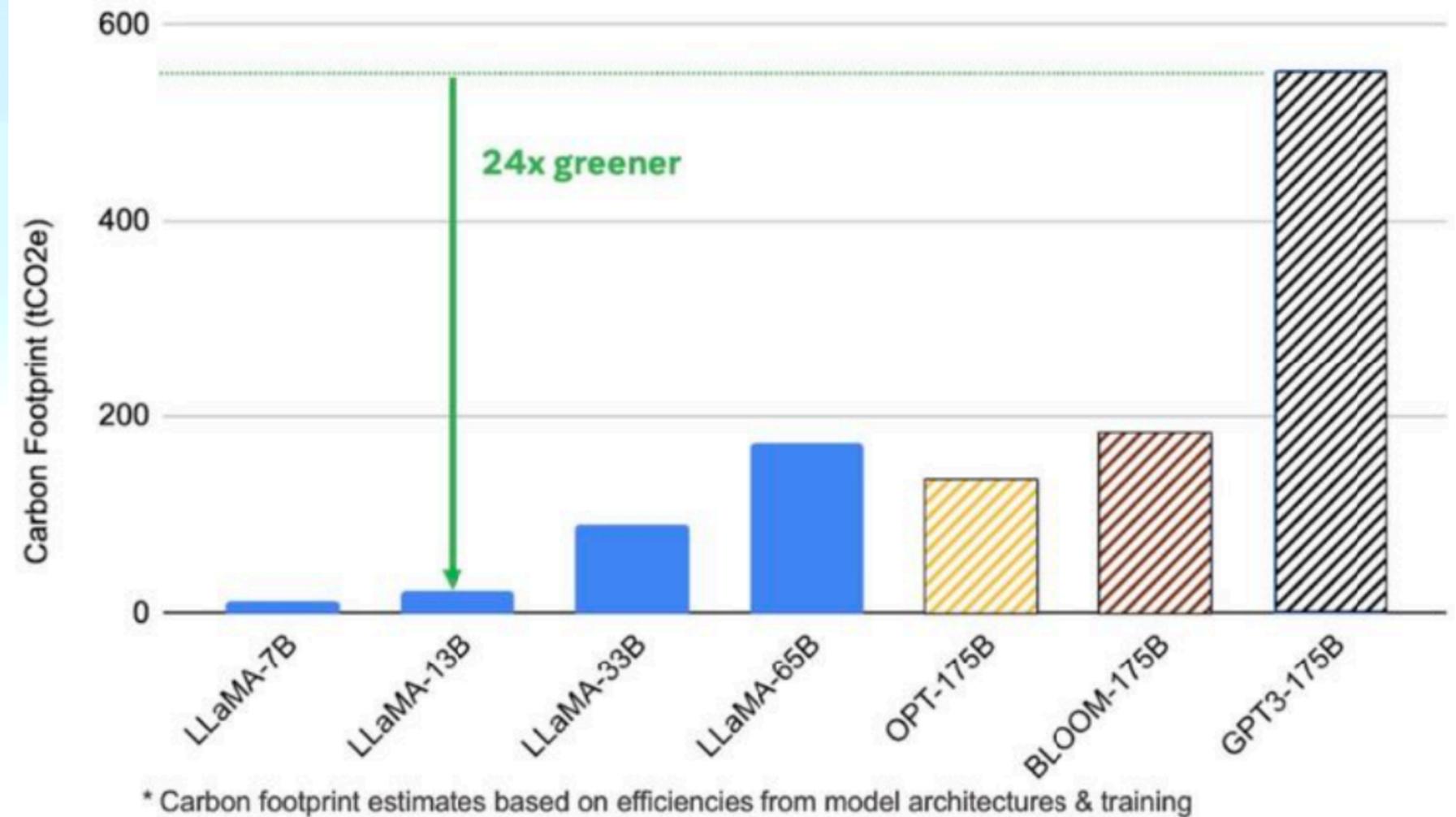
GPT3, Dalle,...,LaMDA 100-200 miliardi di parametri

Modelli di dimensioni enormi addestrati su “corpus” giganteschi, tecnica dello starter byte, necessita1 di prompt engineering

Instruct GPT (1-2 miliardi), ChatGPT (20 miliardi ??)

Addestramento rinforzato dai feedback umani (RJHF)

Operational Carbon Footprint of LLM Training



da rapporto Web di META Corp, Apr. 2023

Gli aspetti etici

L'introduzione dell' IA solleva preoccupazioni etiche

SPIEGABILITÀ. (Explanability) È necessario essere in grado di tracciare una complessa catena di sistemi algoritmici e processi di dati per scoprire il "perché" in caso di risultati non voluti o dannosi

RESPONSABILITÀ. È uno degli aspetti più critici perché non è facile definire le responsabilità quando le decisioni prese dai sistemi di intelligenza artificiale hanno conseguenze inattese o dannose: perdita di capitale, aspetti di salute, diritti d'autore, incidenti mortali, violazione della privacy ecc.

EQUITÀ. Accesso possibile a tutti in modo che tutti abbiano la disponibilità delle potenzialità

Nessun pregiudizio in termini di razza, genere o etnia.

Il pregiudizio dell'apprendimento automatico si ha quando un algoritmo produce risultati sistematicamente influenzati a causa di presupposti errati

qualità, dall'obiettività e dalla quantità dei dati di addestramento

USO IMPROPRIO. introdurre in fase di progettazione la riduzione al minimo dei rischi

La ricerca molecolare e genetica può portare ad usi con conseguenze inaspettate o volutamente negative. Materiali nuovi che si rivelino dannosi per l'ambiente nel tempo, manipolazioni genetiche oltre la prevenzione di malattie, nuove armi

Le leggi di Asimov

dalla fantascienza anni '80 alla realtà di oggi

- **Un robot non può recar danno a un essere umano** né può permettere che, a causa del suo mancato intervento, un essere umano riceva danno.
- **Un robot deve obbedire agli ordini impartiti** dagli esseri umani, purché tali ordini non vadano in contrasto alla Prima Legge.
- **Un robot deve proteggere la propria esistenza**, purché la salvaguardia di essa non contrasti con la Prima o con la Seconda Legge.

The ASILOMAR principles

Proposto da
specialisti di varie
discipline del
M.I.T.

Asilomar AI principles

RESEARCH

1. Research goal
2. Research funding
3. Science-policy link
4. Research culture
5. Race avoidance

ETHICS AND VALUES

6. Safety
7. Failure transparency
8. Judicial transparency
9. Responsibility
10. Value alignment
11. Human values
12. Personal privacy
13. Liberty and privacy
14. Shared benefit
15. Shared prosperity
16. Human control
17. Non-subversion
18. AI arms race

LONGER-TERM ISSUES

19. Capability caution
20. Importance
21. Risks
22. Recursive self-improvement
23. Common good



Rome Call for AI Ethics

From February 2020 to November 2022

Febbraio 2020 - Pontificia Accademia per la Vita, IBM, FAO Microsoft e dall'Associazione italiana Ministero dell'Innovazione

impegno in materia di etica, diritti e istruzione,

obiettivo di promuovere un approccio etico alla progettazione, allo sviluppo e all'implementazione dell'IA

3 aree: Educazione Etica e Diritti

6 Principi: Trasparenza, Inclusione, Responsabilità, Imparzialità, Affidabilità, Sicurezza e Privacy

Nov 2022 workshop internazionale “Convergere sulla persona. Tecnologie emergenti per il bene comune” con rappresentanti di grandi multinazionali dell’IT e rappresentanti delle 3 religioni monoteiste

> *Mantenere l’essere umano al centro dell’evoluzione tecnologica, e non una tecnologia fine a se stessa*



COMMISSIONE EUROPEA

Bruxelles, 21.4.2021

COM(2021) 206 final

2021/0106(COD)

Proposta di

REGOLAMENTO DEL PARLAMENTO EUROPEO E DEL CONSIGLIO

CHE STABILISCE REGOLE ARMONIZZATE SULL'INTELLIGENZA ARTIFICIALE (LEGGE SULL'INTELLIGENZA ARTIFICIALE) E MODIFICA ALCUNI ATTI LEGISLATIVI DELL'UNIONE

{SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final}

RELAZIONE

1. CONTESTO DELLA PROPOSTA

1.1. Motivi e obiettivi della proposta

La presente relazione accompagna la proposta di regolamento che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale). Con il termine intelligenza artificiale (IA) si indica una famiglia di tecnologie in rapida evoluzione in grado di apportare una vasta gamma di benefici economici e sociali in tutto lo spettro delle attività industriali e sociali. L'uso dell'intelligenza artificiale, garantendo un miglioramento delle previsioni, l'ottimizzazione delle operazioni e dell'assegnazione delle risorse e la personalizzazione dell'erogazione di servizi, può contribuire al conseguimento di risultati vantaggiosi dal punto di vista sociale e ambientale nonché fornire vantaggi competitivi fondamentali alle imprese e all'economia europea. Tale azione è particolarmente necessaria in settori ad alto impatto, tra i quali figurano quelli dei cambiamenti climatici, dell'ambiente e della sanità, il settore pubblico, la finanza, la mobilità, gli affari interni e l'agricoltura. Tuttavia gli stessi elementi e le stesse tecniche che alimentano i

Aree di discussione

Prepariamoci oggi per affrontare la rivoluzione

Siamo all'inizio di un viaggio, con tutte le aspettative e i rischi tipici dei viaggi

- L'Intelligenza Artificiale è uno strumento potente. Ognuno per la propria attività deve valutare gli impatti e come affrontare il cambiamento.
- L'etica ha un ruolo importante per un'evoluzione verso una "Good AI" in modo pro-attivo. Non solo normative ma attitudine diffusa.
- Formazione come primo elemento etico di accesso consapevole.
- Una "Good AI depends on the models, boundaries and governance" (intervista a IBM's exec Mark Johnsons del 16 Giugno a Federal News")
 - Ogni attività o impresa dovrebbe identificare gli impatti attuali e futuri dell'introduzione di Intelligenza Artificiale
 - Può diventare necessario avere un responsabile dedicato
 - Minimizzare le esposizioni legali e di relazione con altre entità





GRAZIE



L'intelligenza artificiale permea molti aspetti della vita quotidiana e professionale. Dare qualche minimo elemento ai vari professionisti per cercare di rendere visibile e comprensibile l'azione di questi strumenti e poterli gestire per non essere estromessi dal processo decisionale è l'obiettivo di questa pubblicazione, che non ha la pretesa di essere esaustiva, ma cerca di affrontare una prima introduzione all'Intelligenza Artificiale per chi, già ora o nel breve termine, avrà a che fare con soluzioni che la utilizzano.



Roberto Magnani, Laureato in Ingegneria Elettronica, sposato con un figlio, ha sviluppato la sua carriera nei laboratori di multinazionali in Italia, Francia, USA, Spagna e Irlanda. Nell'ultimo decennio è stato il leader per l'Europa dei servizi via Web del Public Cloud di una grande multinazionale in un Campus Tecnologico di Dublino per poi assumere la leadership dello sviluppo di progetti digitali per Healthcare/Life Science per EMEA, con l'utilizzo di Intelligenza Artificiale. Dal 2022 svolge

attività di consulenza indipendente e recentemente ricopre il ruolo di consigliere nell'italiana AEIT (Associazione Elettronica Elettrotecnica e Informatica - ente morale dal 1897 per la promozione dell'innovazione in campo delle tecnologie) focalizzandosi sugli aspetti etici e normativi per l'Intelligenza Artificiale e l'introduzione del Quantum Computing nell'industria. È autore di articoli e interventi in Italia e all'estero sugli stessi argomenti.

€ 15,90



eb

ROBERTO MAGNANI

INTELLIGENZA ARTIFICIALE PER LE PROFESSIONI
EDIZIONE PER NON SPECIALISTI DI INFORMATICA

ROBERTO MAGNANI

INTELLIGENZA ARTIFICIALE PER LE PROFESSIONI

EDIZIONE PER NON SPECIALISTI DI INFORMATICA



etabeta