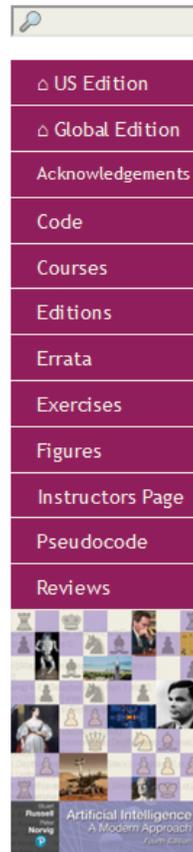


# Introduzione all'Intelligenza Artificiale (IA)

# Disclaimer (1)



## Artificial Intelligence: A Modern Approach, 4th US ed.

by Stuart Russell and Peter Norvig

The [authoritative, most-used](#) AI textbook, adopted by over [1500](#) schools.

Table of Contents for the US Edition (or see the [Global Edition](#))

[Preface \(pdf\)](#); [Contents with subsections](#)

### I Artificial Intelligence

- 1 Introduction ... 1
- 2 Intelligent Agents ... 36

### II Problem-solving

- 3 Solving Problems by Searching ... 63
- 4 Search in Complex Environments ... 110
- 5 Adversarial Search and Games ... 146
- 6 Constraint Satisfaction Problems ... 180

### III Knowledge, reasoning, and planning

- 7 Logical Agents ... 208
- 8 First-Order Logic ... 251
- 9 Inference in First-Order Logic ... 280
- 10 Knowledge Representation ... 314
- 11 Automated Planning ... 344

### IV Uncertain knowledge and reasoning

- 12 Quantifying Uncertainty ... 385
- 13 Probabilistic Reasoning ... 412
- 14 Probabilistic Reasoning over Time ... 461
- 15 Probabilistic Programming ... 500
- 16 Making Simple Decisions ... 528
- 17 Making Complex Decisions ... 562
- 18 Multiagent Decision Making ... 599

### V Machine Learning

- 19 Learning from Examples ... 651
- 20 Learning Probabilistic Models ... 721
- 21 Deep Learning ... 750
- 22 Reinforcement Learning ... 789

### VI Communicating, perceiving, and acting

- 23 Natural Language Processing ... 823
- 24 Deep Learning for Natural Language Processing ... 856
- 25 Computer Vision ... 881
- 26 Robotics ... 925

### VII Conclusions

- 27 Philosophy, Ethics, and Safety of AI ... 981
- 28 The Future of AI ... 1012
- Appendix A: Mathematical Background ... 1023
- Appendix B: Notes on Languages and Algorithms ... 1030
- Bibliography ... 1033 ([pdf](#) and [bib data](#))
- Index ... 1069 ([pdf](#))

[Exercises \(website\)](#)

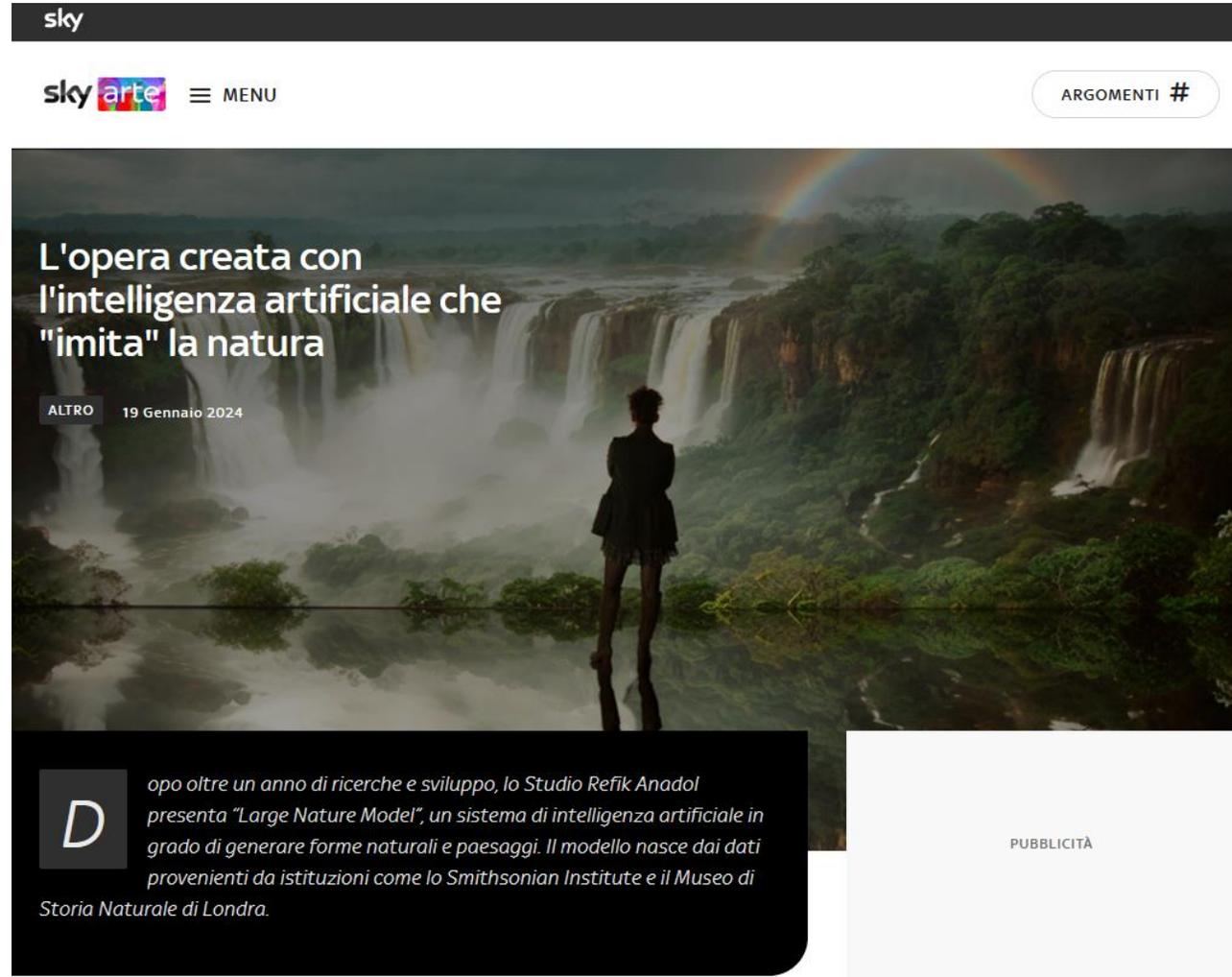
[Figures \(pdf\)](#)

[Code \(website\)](#); [Pseudocode \(pdf\)](#)

Covers: [US](#), [Global](#)

# Perché l'Intelligenza Artificiale?

# Perché l'IA?



The screenshot shows a news article on the Sky Arte website. The main image is a digital artwork of a person standing on a rocky outcrop, looking out over a vast, lush landscape with multiple waterfalls and a rainbow in the sky. The article title is "L'opera creata con l'intelligenza artificiale che 'imita' la natura". Below the title, it says "ALTRO" and "19 Gennaio 2024". A text box at the bottom left contains the start of the article's introduction, and a "PUBBLICITÀ" label is visible on the right side of the article area.

sky

sky arte MENU

ARGOMENTI #

## L'opera creata con l'intelligenza artificiale che "imita" la natura

ALTRO 19 Gennaio 2024

**D**opo oltre un anno di ricerche e sviluppo, lo Studio Refik Anadol presenta "Large Nature Model", un sistema di intelligenza artificiale in grado di generare forme naturali e paesaggi. Il modello nasce dai dati provenienti da istituzioni come lo Smithsonian Institute e il Museo di Storia Naturale di Londra.

PUBBLICITÀ

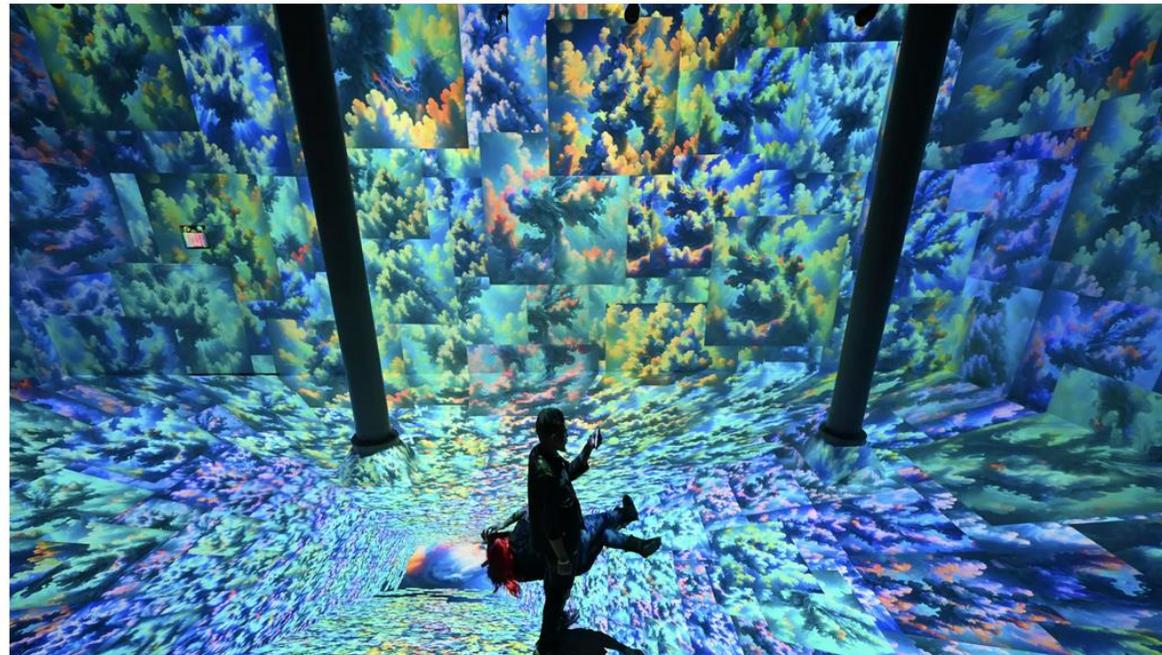
<https://arte.sky.it/news/mostre-milano-stagione-2024-2025>

# Perché l'IA?

ALESSANDRO PATELLA ECONOMIA 23.10.2024

## Contro l'intelligenza artificiale che si "allena" su opere d'arte firmano 6.500 artisti

Appello per bloccare l'addestramento degli algoritmi di AI generativa su opere protette da diritto d'autore



<https://www.wired.it/article/intelligenza-artificiale-generativa-appello-artisti-contro-copyright/>

# Perché l'IA?

## Intelligenza artificiale generativa, domanda in crescita del 246%

Il riferimento è presente in 10.795 annunci su LinkedIn, analizzati da Gi group holding e Microsoft Italia. Più del 10% lo indica nel job title

di Cristina Casadei

30 ottobre 2024



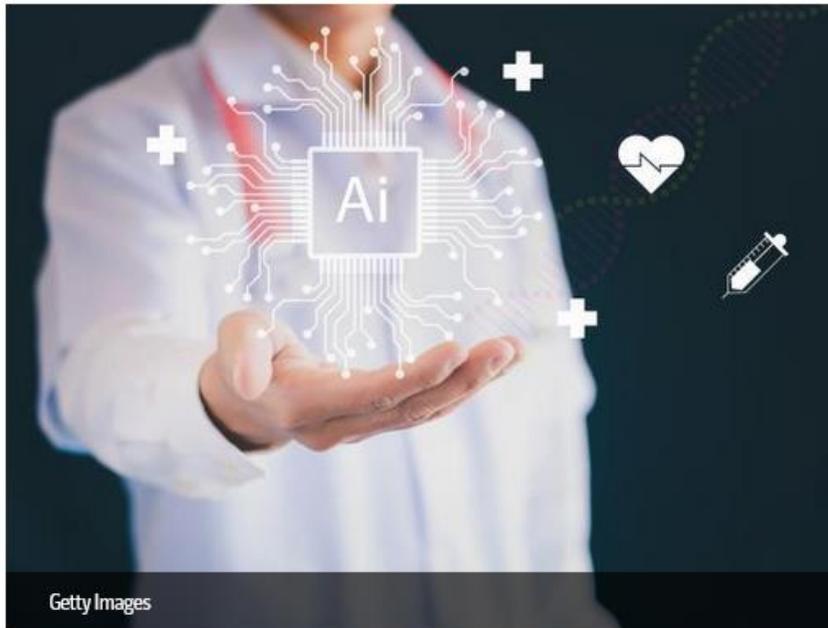
[https://www.ilsole24ore.com/art/intelligenza-artificiale-generativa-domanda-crescita-246percento-AGVscyo?refresh\\_ce=1](https://www.ilsole24ore.com/art/intelligenza-artificiale-generativa-domanda-crescita-246percento-AGVscyo?refresh_ce=1)

# Perché l'IA?

## **Intelligenza artificiale in medicina: da quella generativa, alle reti neurali e ai chatbot. Come viene (sempre più) utilizzata**

di Ruggiero Corcella

Gli strumenti di AI si applicano alla diagnosi, prognosi e cura del paziente. Quello che ne deriva è una medicina di precisione, personalizzata e cucita su misura. Lo studio di NetConsulting Cube: «Crescita esponenziale»



**DIZIONARIO DELLA SALUTE**

Cerca il tuo organo/patologia

**CORRIERE TV**

Attiva audio

LA PRIMA CORNEA IBRIDA ARTIFICIALE SVILUPPATA IN ITALIA

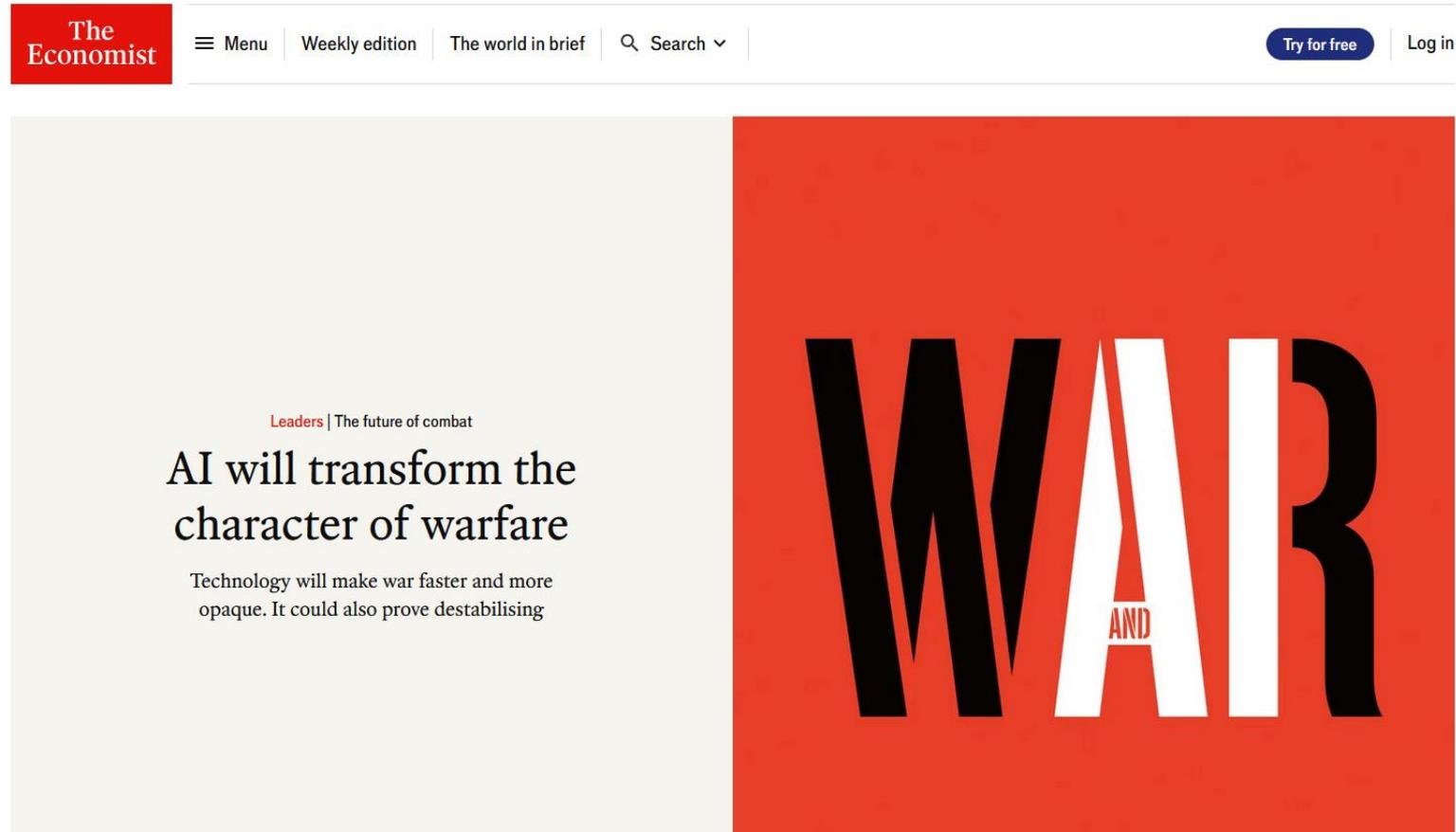
**Cornea artificiale ibrida «italiana»: si può tornare a vedere**

A Forlì i primi tre interventi con il nuovo dispositivo hanno dato risultati incoraggianti. La tecnica può offrire una soluzione ai casi in cui il trapianto di cornea da donatore non viene tollerato

EDITORIALI & COMMENTI

[https://www.corriere.it/salute/ehealth/24\\_novembre\\_06/intelligenza-artificiale-in-medicina-da-quella-generativa-alle-reti-neurali-e-ai-chatbot-come-viene-sempre-piu-utilizzata-e788084d-d14b-4cee-bdaa-154f2afe4xlk.shtml](https://www.corriere.it/salute/ehealth/24_novembre_06/intelligenza-artificiale-in-medicina-da-quella-generativa-alle-reti-neurali-e-ai-chatbot-come-viene-sempre-piu-utilizzata-e788084d-d14b-4cee-bdaa-154f2afe4xlk.shtml)

# Perché l'IA?



<https://www.economist.com/leaders/2024/06/20/war-and-ai>

# Applicazioni (quotidiane o quasi)

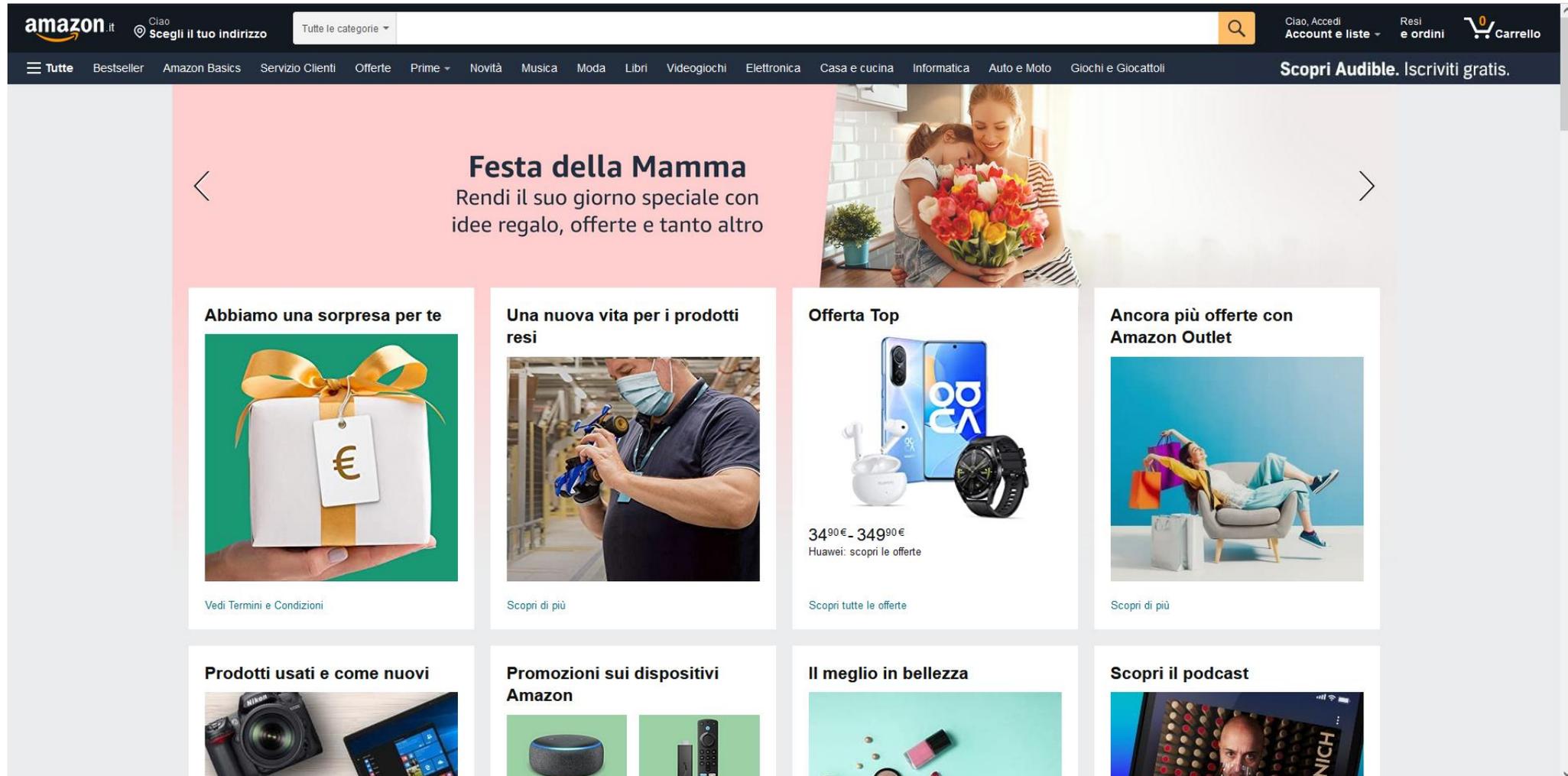
# Recommender Systems



**Recommender system:** sistema di filtraggio dei contenuti che genera raccomandazioni personalizzate, specifiche per l'utente

Esempi?

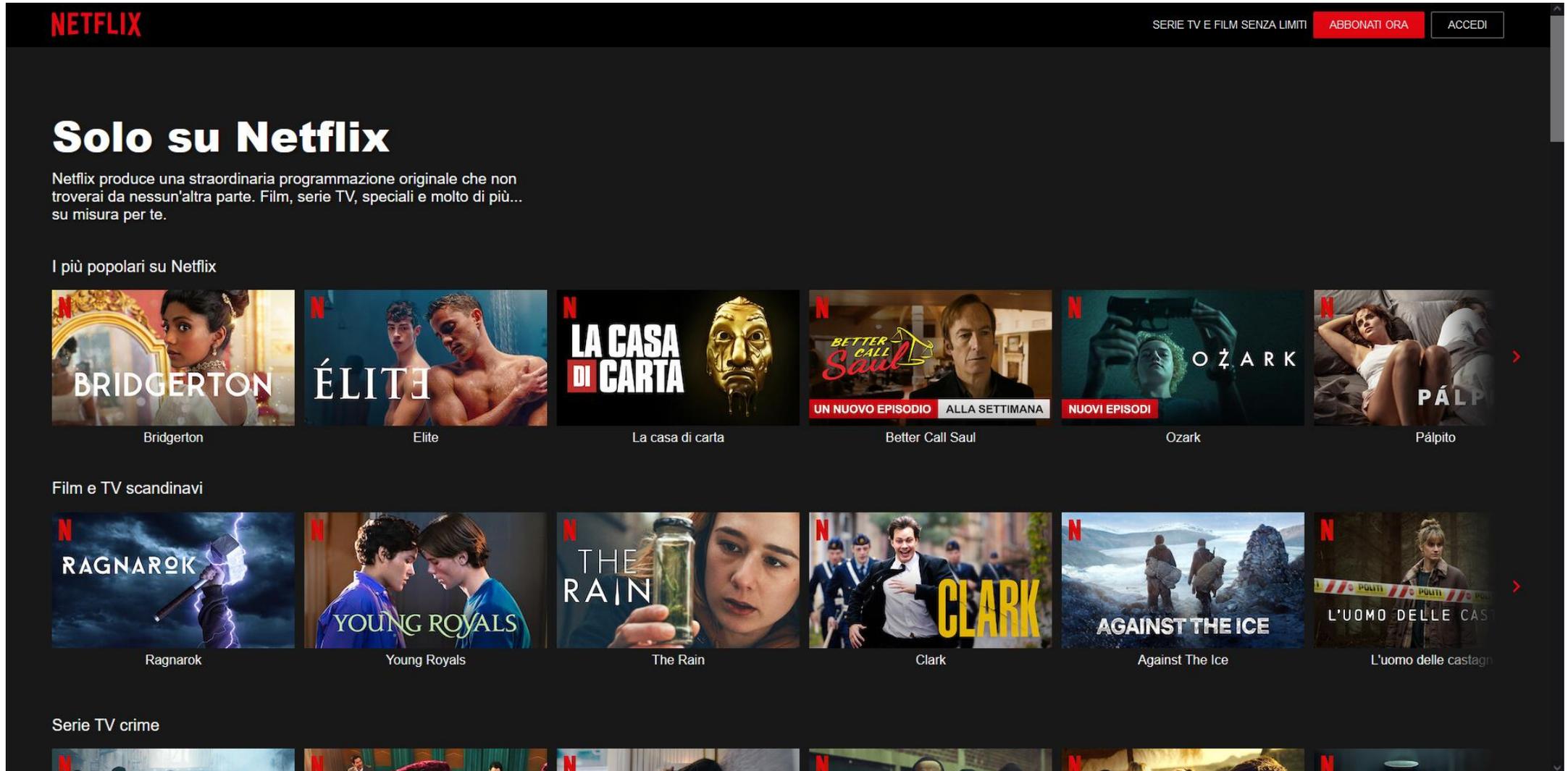
# Recommender Systems



The screenshot shows the Amazon Italy homepage with a navigation bar at the top. The main banner is for 'Festa della Mamma' (Mother's Day), with the text 'Rendi il suo giorno speciale con idee regalo, offerte e tanto altro' (Make her special day special with gift ideas, offers and much more). Below the banner are several promotional tiles:

- Abbiamo una sorpresa per te**: A white gift box with a gold ribbon and a tag with a Euro symbol (€). Below the image is the link 'Vedi Termini e Condizioni'.
- Una nuova vita per i prodotti resi**: A man in a blue shirt and mask working with a drone. Below the image is the link 'Scopri di più'.
- Offerta Top**: A Huawei smartphone, a smartwatch, and a smart speaker. The price is shown as '34<sup>90</sup>€ - 349<sup>90</sup>€'. Below the image is the link 'Scopri tutte le offerte'.
- Ancora più offerte con Amazon Outlet**: A woman sitting on a sofa with shopping bags. Below the image is the link 'Scopri di più'.
- Prodotti usati e come nuovi**: A Nikon camera and a tablet.
- Promozioni sui dispositivi Amazon**: An Amazon Echo smart speaker and a remote control.
- Il meglio in bellezza**: A collection of beauty products.
- Scopri il podcast**: A smartphone displaying a podcast cover.

# Recommender Systems



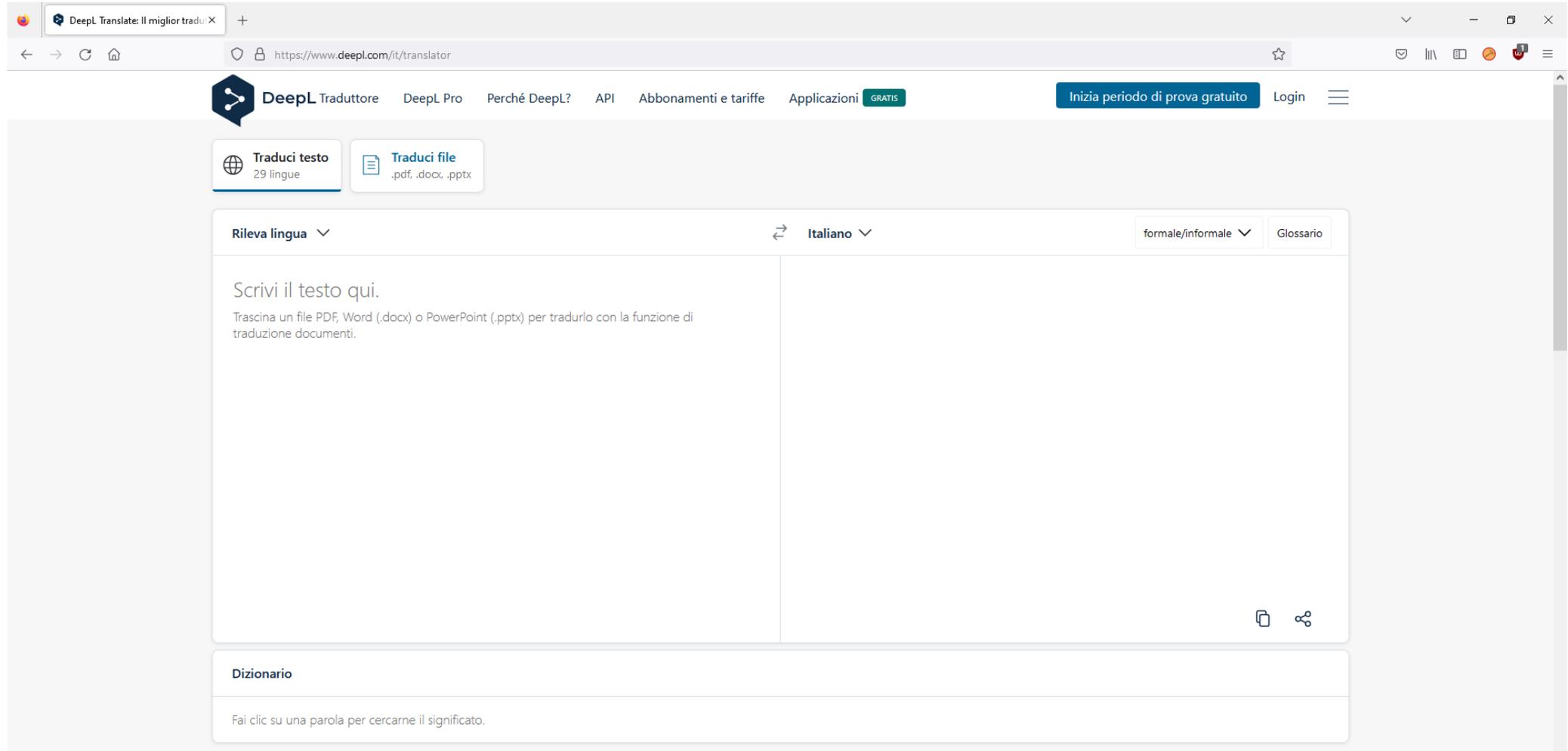
The screenshot shows the Netflix homepage with the following elements:

- NETFLIX** logo in the top left.
- Navigation links: **SERIE TV E FILM SENZA LIMITI**, **ABBONATI ORA**, and **ACCEDI**.
- Solo su Netflix** section with the text: "Netflix produce una straordinaria programmazione originale che non troverai da nessun'altra parte. Film, serie TV, speciali e molto di più... su misura per te."
- I più popolari su Netflix** section featuring a row of six titles: **BRIDGERTON**, **ÉLITE**, **LA CASA DI CARTA**, **BETTER CALL SAUL** (with a banner: "UN NUOVO EPISODIO ALLA SETTIMANA"), **OZARK** (with a banner: "NUOVI EPISODI"), and **PÁLPITO**.
- Film e TV scandinavi** section featuring a row of six titles: **RAGNAROK**, **YOUNG ROYALS**, **THE RAIN**, **CLARK**, **AGAINST THE ICE**, and **L'UOMO DELLE CASTAGNE**.
- Serie TV crime** section with a row of six titles partially visible.

# Riconoscimento facciale



# Traduzione automatica

A screenshot of the DeepL Translator website interface. The browser address bar shows 'https://www.deepl.com/it/translator'. The page header includes the DeepL logo and navigation links: 'DeepL Traduttore', 'DeepL Pro', 'Perché DeepL?', 'API', 'Abbonamenti e tariffe', 'Applicazioni', and a 'GRATIS' badge. A blue button for 'Inizia periodo di prova gratuito' and a 'Login' link are also present. The main content area has two tabs: 'Traduci testo' (29 lingue) and 'Traduci file' (.pdf, .docx, .pptx). The 'Traduci testo' tab is active, showing a language selection dropdown set to 'Italiano', a 'formale/informale' dropdown, and a 'Glossario' link. The main text area contains the instruction 'Scrivi il testo qui.' and a note about file uploads. A 'Dizionario' section is visible at the bottom with the text 'Fai clic su una parola per cercarne il significato.'

<https://www.deepl.com/it/translator>

# Politica

Può l'IA favorire una fazione durante la campagna elettorale?



# Cambridge Analytica (~2016)



- Azienda specializzata nella fornitura di servizi per pubblicità mirata;
- Un'altra azienda, "thisisyourdigitallife", raccoglie (legittimamente) i dati di milioni di utenti;
- Compra i dati da "thisisyourdigitallife" (qui avviene l'illecito);
- Usa quei dati per fornire pubblicità mirata in favore di uno dei candidati alle presidenziali 2016.

## Altre applicazioni

- marketing e customer care (es. sentiment analysis)
- medicina e biologia (bioinformatica)
- riconoscimento di immagini
- ...

Alcuni esempi: <https://machinelearningmastery.com/inspirational-applications-deep-learning/>



# Cos'è l'Intelligenza Artificiale?

# Cos'è l'IA?

“The science and engineering of making intelligent machines”

---

John McCarthy (1956)



Trenchard More, John McCarthy, Marvin Minsky, Oliver Selfridge, Ray Solomonoff  
Tecnologie e Processi Digitali 24/25

# Cos'è l'IA?



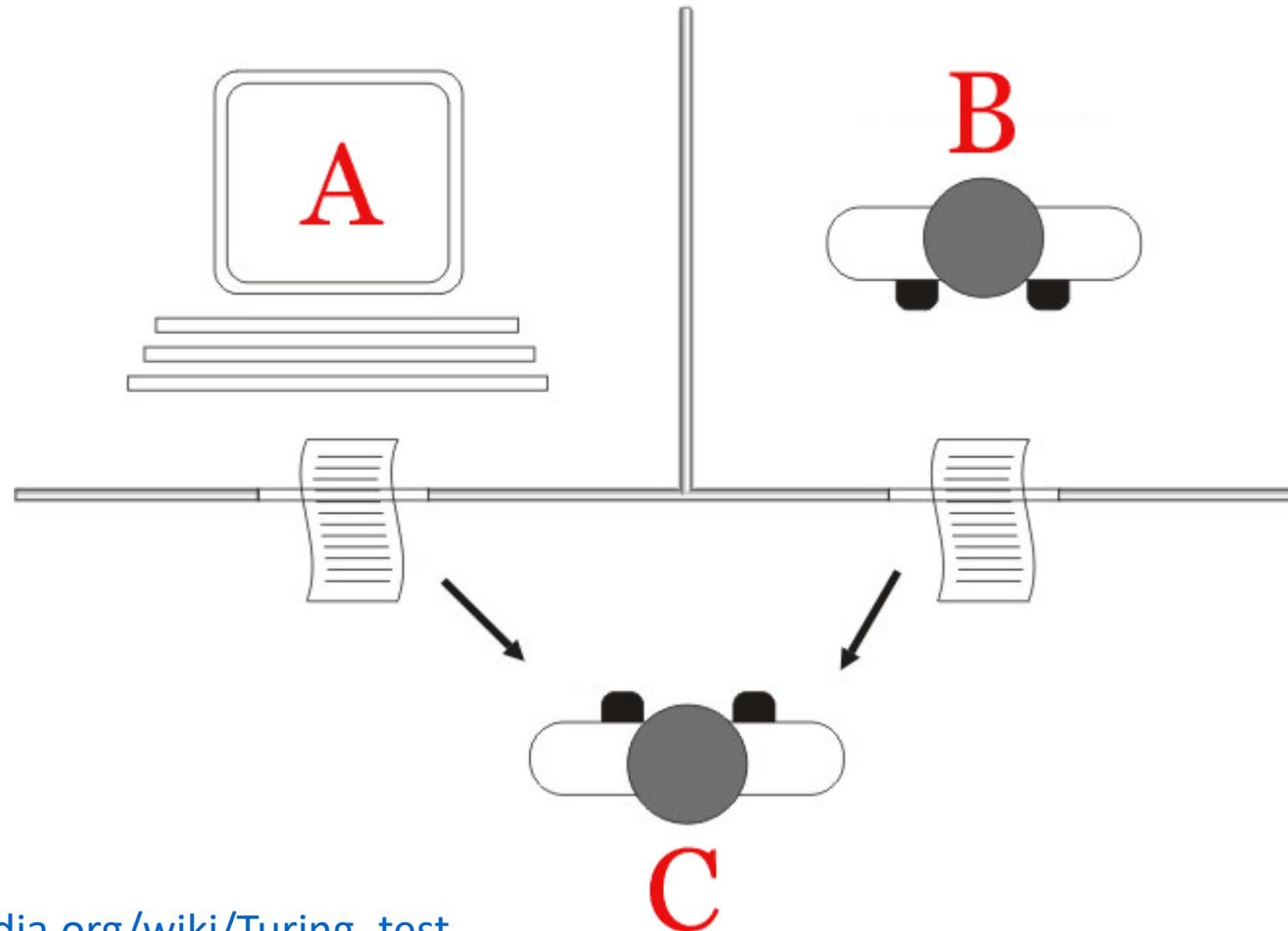
Macchine (sistemi) che pensano/agiscono **come umani?**

OPPURE

Macchine (sistemi) che pensano/agiscono **razionalmente?**

# Macchine che agiscono **come umani**

## Test di Turing



Da [https://en.wikipedia.org/wiki/Turing\\_test](https://en.wikipedia.org/wiki/Turing_test)

# Macchine che pensano **come umani**



**IA Forte/IA generale:** La capacità di una macchina di comprendere o apprendere ogni tipo di compito «intellettuale» che un essere umano è in grado di comprendere o apprendere. In altre parole, una macchina che ha coscienza di sé.

**IA Debole/IA ristretta:** La capacità di una macchina di implementare una parte di «intelligenza» per eseguire un compito preciso. L'intelligenza artificiale con cui abbiamo a che fare quotidianamente è di questo tipo (es. assistente vocale, riconoscimento immagini... tutte le applicazioni viste in precedenza!)

# Chinese room – John Searle (1980)

Esperimento mentale “contro” IA forte

- ipotesi: un computer che si comporta come se comprendesse il cinese;
- ipotesi: il computer passa il test di Turing;
- il computer comprende davvero il cinese o simula di comprenderlo?

Un essere umano senza alcuna conoscenza del cinese, messo in una stanza con un manuale con le opportune istruzioni, potrebbe fare la stessa cosa, senza comprendere davvero ciò che gli viene chiesto

# Macchine che pensano/agiscono **razionalmente**



**unIMC**

**Comportamento razionale:** eseguire un'azione il cui risultato atteso massimizza un certo obiettivo (perlomeno secondo l'informazione disponibile al momento dell'esecuzione).

Un agente intelligente *delibera* un'azione da eseguire (o nuovi fatti da aggiungere alla propria base di conoscenza) mediante

- Deduzione
- Abduzione
- Induzione

# Deduzione

Il processo logico nel quale, date certe premesse e certe regole che ne garantiscono la correttezza, una conclusione consegue come logicamente necessaria. E' lo schema di ragionamento del sillogismo.

## Sillogismo

Ogni uomo è mortale  
Socrate è un uomo

ALLORA

Socrate è mortale

$\forall(x)(influenza(x) \rightarrow febbre(x))$

$\Rightarrow$

$febbre(Socrate)$

$influenza(Socrate)$

# Abduzione

Lo schema di ragionamento tipico della diagnosi

$\forall(x)(influenza(x) \rightarrow febbre(x))$

$febbre(Socrate)$

---

$influenza(Socrate)$

# Induzione

L'apprendimento dalle osservazioni

*influenza(Socrate)*

*febbre(Socrate)*

*influenza(Aristotele)*

*febbre(Aristotele)*

*influenza(Euclide)*

*febbre(Euclide)*

...

---

$$\forall(x)(influenza(x) \rightarrow febbre(x))$$

## IA simbolica vs IA *data-driven*

**IA simbolica:** rappresentazione logico-matematica del mondo. Esempi: ricerca nello spazio degli stati, problemi di ottimizzazione, ecc.

**IA *data-driven*:** apprendimento automatico, senza rappresentazione esplicita delle regole/funzioni apprese

# IA simbolica

Decisioni spiegabili per definizione

Non è semplice trovare la rappresentazione opportuna

Non tutto si può computare tempestivamente...

## IA *data-driven*

Apprendimento autonomo di relazioni sconosciute

Numero di esempi rappresentativo della realtà

Scatola nera/decisioni non spiegabili

## IA – Storicamente...

Forte, fin dai primi anni, in problemi descrivibili come un set di formule logiche/matematiche

- theorem proving e motori di inferenza;
- sistemi esperti;
- logic based knowledge representation systems;
- problem solving e ricerca;
- ...

Questo nonostante i noti problemi di intrattabilità computazionale e decidibilità

# IA – Storicamente (1997)



# IA – Storicamente

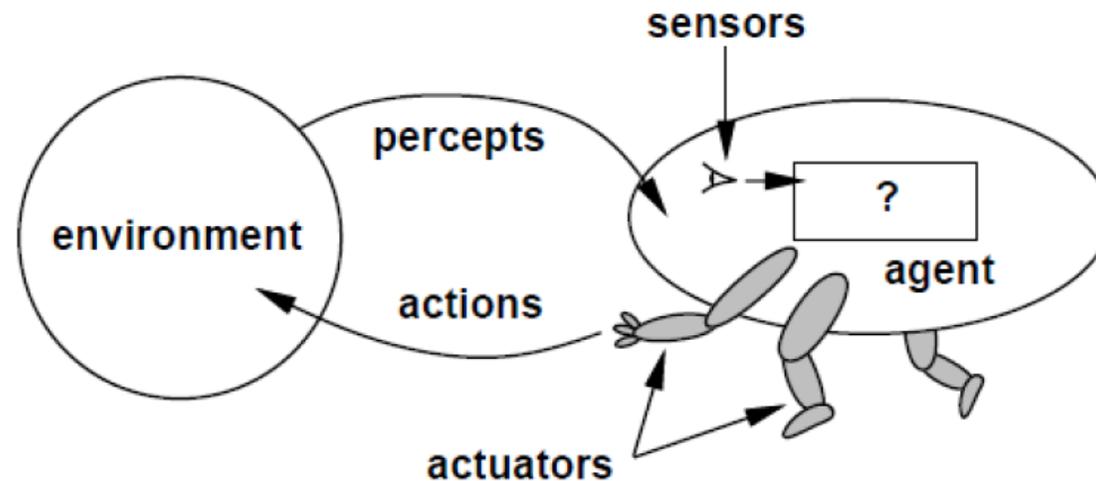
- Forte in problemi «intellettualmente complessi» per gli umani
- Debole in problemi «facili» per gli umani ma complessi da descrivere formalmente
  - riconoscimento del parlato
  - riconoscimento di oggetti in immagini
  - riconoscimento di volti
  - ...

# IA vs «conoscenza informale»

Task apparentemente semplici come riconoscere oggetti richiedono esperienza, formata mediante apprendimento

# Altre definizioni – Agente Intelligente

“An agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through effectors.”



Da **S. Russell** and **P. Norvig**, *Artificial Intelligence: A Modern Approach*, 3rd ed. Prentice Hall, 2009.



# Machine Learning (ML – Apprendimento Automatico)

# Machine Learning

L'apprendimento automatico (machine learning) è l'insieme di metodi per permettere ai computer di apprendere autonomamente dagli esempi. Più precisamente l'apprendimento automatico studia e sviluppa algoritmi per apprendere dai dati, usarli per fare previsioni e scoprire nuova conoscenza.

Da: A. L. Samuel, "*Some studies in machine learning using the game of checkers*". IBM Journal of research and development, vol. 3 (3), pp. 210-229, 1959

# Machine Learning

L'apprendimento automatico (machine learning) è l'insieme di metodi per permettere ai computer di apprendere autonomamente dagli esempi. Più precisamente l'apprendimento automatico studia e sviluppa algoritmi per apprendere dai dati, usarli per fare previsioni e scoprire nuova conoscenza.



Percezione non solo per agire, ma anche per migliorare la capacità di agire in futuro.

# Problemi di machine Learning



- Classificazione
- Regressione
- Clustering
- Riduzione della dimensionalità

# Classificazione

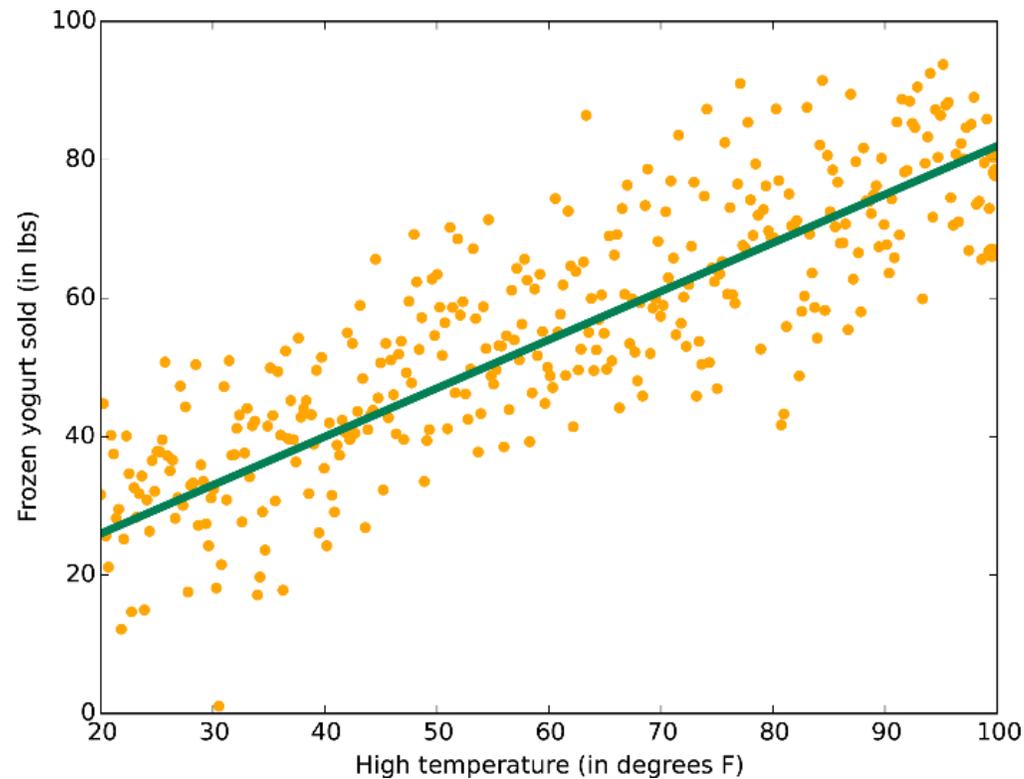
Associa una classe ai dati di input (riconoscimento). Una classe indica un insieme di oggetti aventi proprietà comuni, o, in altre parole, che seguono lo stesso pattern.

Esempi:

- a partire da un vettore di input che rappresenta un'immagine di un volto, riconoscere un soggetto
- riconoscere categorie di oggetti nelle immagini
- riconoscere usi propri e impropri di una carta di credito

# Regressione

Associa un valore ai dati di input. In altre parole, stima una funzione che si avvicini il più possibile ai dati di input. Può essere usata per fare «previsioni»



Da <https://docs.microsoft.com/it-it/azure/machine-learning/studio/algorithm-choice>

# Clustering

Suddivide in gruppi (cluster) i campioni di input con caratteristiche simili. Al contrario della classificazione, quindi, le classi non sono note a-priori.

Esempio:

- profilazione di utenti per il suggerimento di prodotti/acquisti (recommender system)

# Riduzione della dimensionalità

Mapping dei campioni di input in uno spazio a dimensionalità più bassa. In altre parole riduce la cardinalità di ogni campione di input.

Ovviamente c'è una perdita di informazione, ma:

- l'obiettivo è solitamente rendere trattabile un problema;
- mantiene le informazioni «importanti», scartando quelle superflue o ridondanti;
- utile anche per la visualizzazione di dati

# Machine learning «convenzionale»



- conoscenza di dominio per la scelta delle feature da estrarre
- dati «raw» da trasformare nello spazio delle feature
- dati trasformati in input per la classificazione o la scoperta di pattern

# Deep Learning (DL)

# Representation Learning



Può un «agente intelligente» apprendere autonomamente quali feature utilizzare per i problemi di machine learning, a partire da dati «raw»?

**Representation Learning:** insieme di tecniche per passare alle macchine dati «raw» e permettere di «apprendere» la rappresentazione migliore per i task di machine learning da eseguire

# Deep Learning

“Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics.”

Y. LeCun, Y. Bengio, and G. Hinton, “*Deep learning*,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015 (<https://www.nature.com/articles/nature14539>)

# Deep Learning – Timeline

- **1943:** McCulloch e Pitts introducono il primo modello logico-matematico di funzionamento dei neuroni del cervello umano
- **1958:** Roseblatt presenta il percettrone (“perceptron”), algoritmo di apprendimento automatico a due strati (ingresso/uscita)
- **1969:** Minsky e Paperts, nel loro libro “Perceptron”, analizzano alcuni limiti del percettrone
- **1986:** introduzione della backpropagation da parte di Rumelhart, Hinton e Williams.

# Deep Learning – Timeline



- **1989:** Lecun sviluppa reti con connessioni locali e pesi condivisi (Convolutional Neural Network) per il riconoscimento automatico dei caratteri scritti a mano
- **1998:** Lecun e Bengio ottengono un errore inferiore all'1% sul dataset MNIST (30.000 modelli di cifre scritte a mano) con la Lenet-5, una Convolutional Neural Network composta da 7 strati
- **2005:** D. Steinkraus, I. Buck and P. Y. Simard, "*Using GPUs for machine learning algorithms*," Eighth International Conference on Document Analysis and Recognition (ICDAR'05), 2005, pp. 1115-1120 Vol. 2. doi: 10.1109/ICDAR.2005.2513

# Deep Learning – Timeline



- **2012:** nell'annuale "Imagenet Large Scale Visual Recognition Challenge" (ILSVCR), l'approccio proposto da *Krizhevsky, Sutskever e Hinton*\* basato su Convolutional Neural Network ottiene un top-5 test error del 15.3% (il secondo classificato ottiene 26.2%)

---

\*Krizhevsky, A., Sutskever, I. and Hinton, G. *ImageNet classification with deep convolutional neural networks*. In Proc. Advances in Neural Information Processing Systems 25 1090-1098 (2012)

# ILSVCR 2012



- obiettivo: identificare gli oggetti presenti nelle immagini
- training set: 1,2M immagini, 1K categorie di oggetti annotati
- validation e test set: 150K immagini, 1K categorie di oggetti da riconoscere
- task: classificazione (elencare i 5 oggetti riconosciuti con la più alta confidenza)

# Deep Learning – Timeline



- **Marzo 2013:** Google acquisisce “DNNResearch Inc.” di Geoffrey Hinton (e lo assume...)
- **Dicembre 2013:** Yann Lecun diventa direttore di “Facebook AI Research”
- **Gennaio 2014:** Google acquisisce DeepMind
- **Giugno 2014:** DeepFace di Facebook ottiene il 97.35% di accuratezza nel riconoscimento facciale, sul dataset LFW

# Deep Learning – Timeline



- **Dicembre 2015:** Il team del “Microsoft Research Laboratory” vince l’ILSVRC 2015
- **Gennaio 2016:** Microsoft rilascia i sorgenti del proprio “Computational Network Toolkit” CNTK
- **Marzo 2016:** AlphaGo batte Lee Sedol
- **Gennaio 2017:** Microsoft acquisisce Maluuba e “recluta” Yoshua Bengio come consigliere
- ...

# Premio Turing 2018

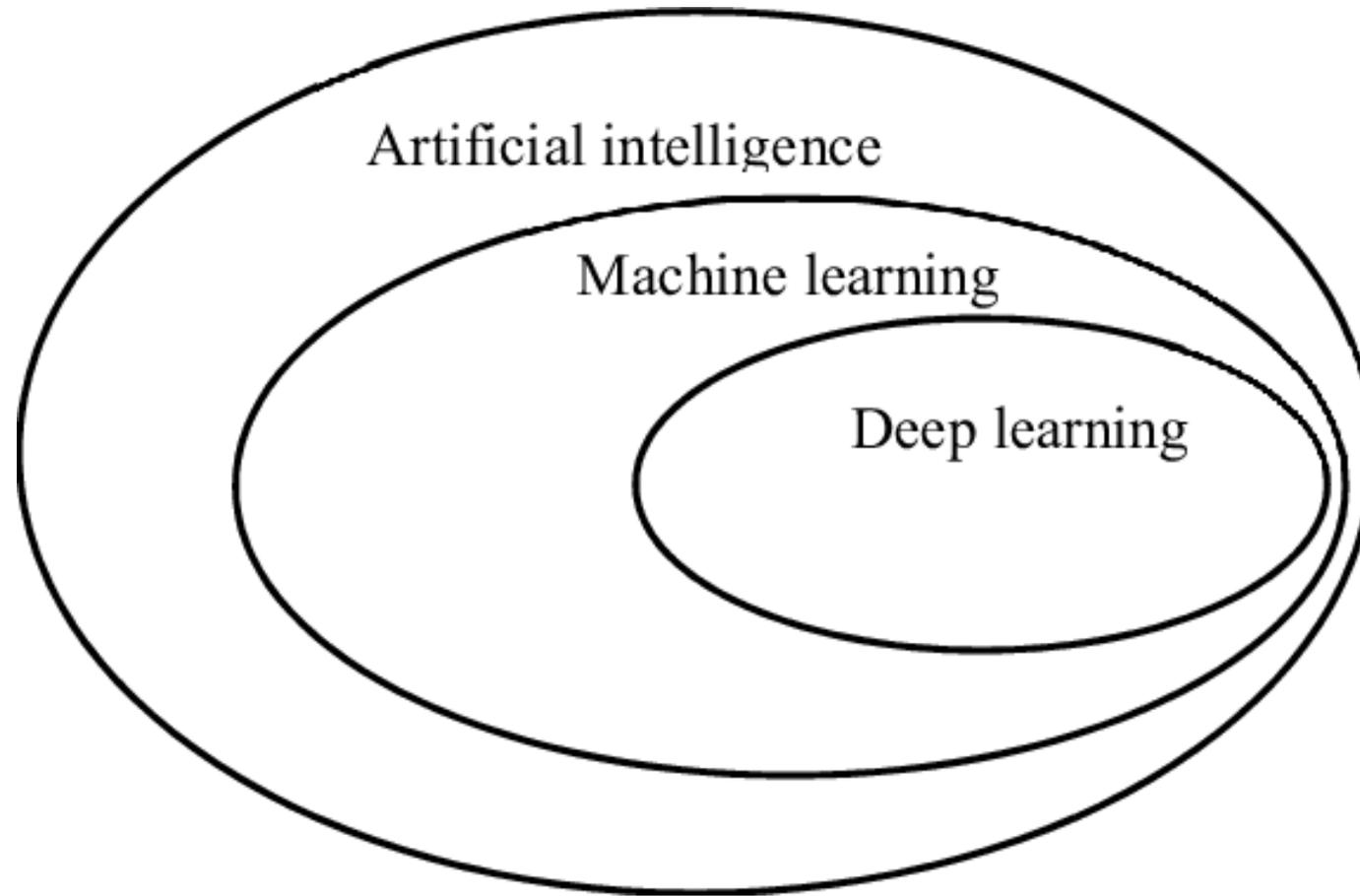
Yann LeCun, Geoffrey Hinton e Yoshua Bengio vengono insigniti del premio Turing 2018



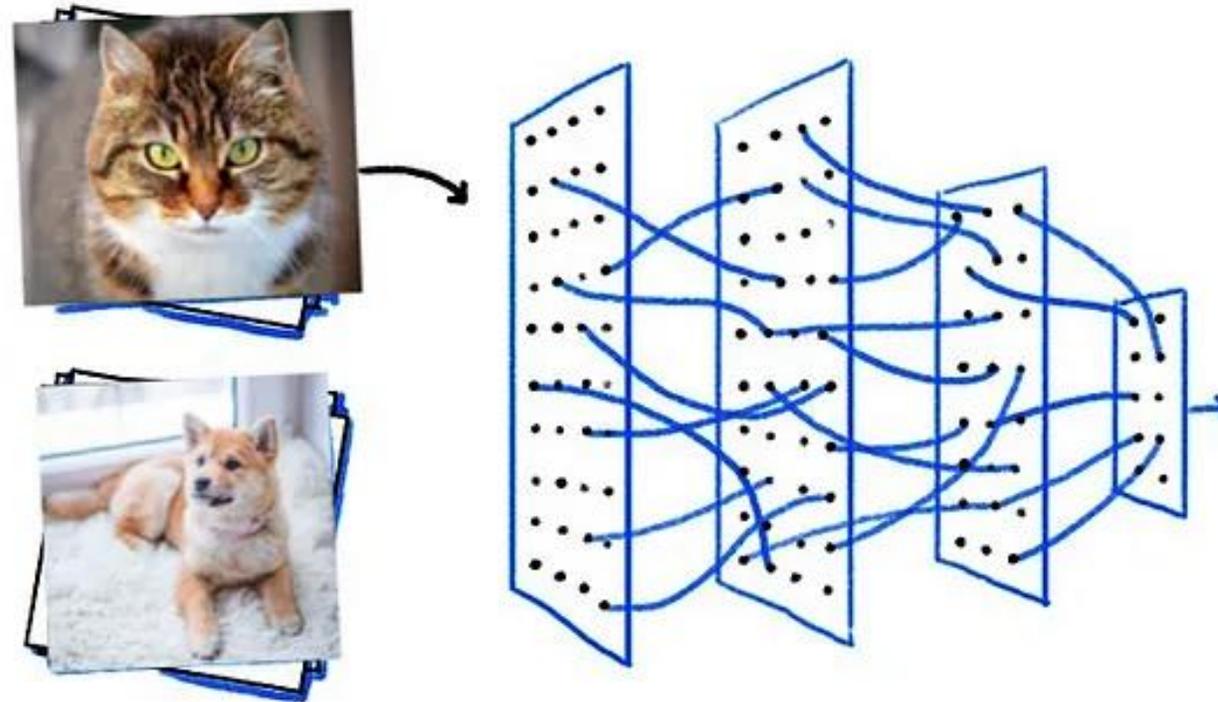
# Nobel per la fisica 2024



# IA, ML e DL



# Artificial Neural Networks



# Reti Neurali (Neural Networks)

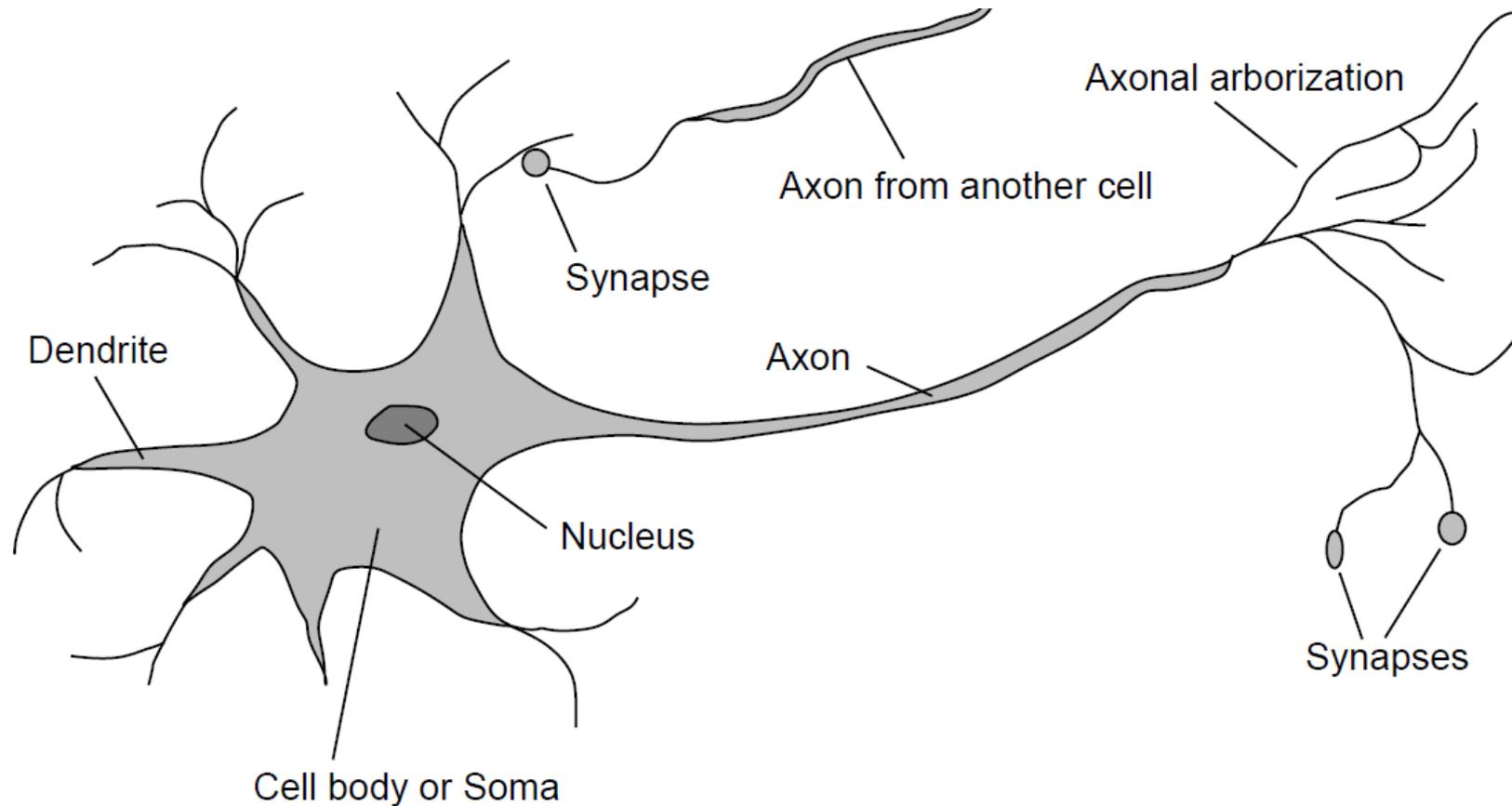
# Riferimenti



- S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 3rd ed. Prentice Hall, 2009.
- M. Nielsen, Neural Networks and Deep Learning. Determination Press, 2015. <http://neuralnetworksanddeeplearning.com>

# Neurone

L'unità fondamentale di tutto il tessuto nervosa, dove avviene l'elaborazione dell'informazione.



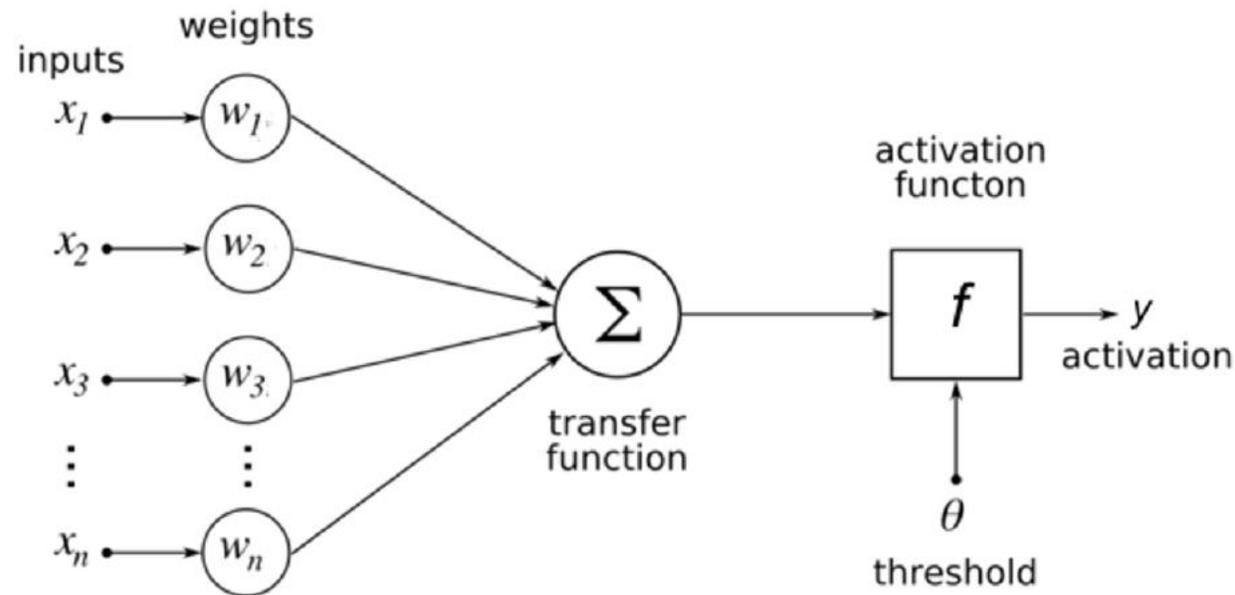
# Neurone

- Sostanze che fungono da trasmettitori chimici vengono rilasciate dalle sinapsi (giunzioni tra neuroni) ed entrano nei dendriti (ingressi del neurone), innalzando o abbassando il potenziale elettrico di un neurone;
- quando il potenziale raggiunge una soglia, il potenziale di attivazione del neurone, un impulso elettrico viene inviato lungo l'assone (l'output del neurone);
- l'impulso elettrico corre lungo le ramificazioni dell'assone, connesse tramite sinapsi ai dendriti di altri neuroni, provocando il rilascio dei trasmettitori nei corpi di altre cellule.

# Neurone Artificiale

Modello matematico (semplificato) del neurone biologico

$$y = f\left(\sum_{i=1}^n w_i x_i - \theta\right)$$



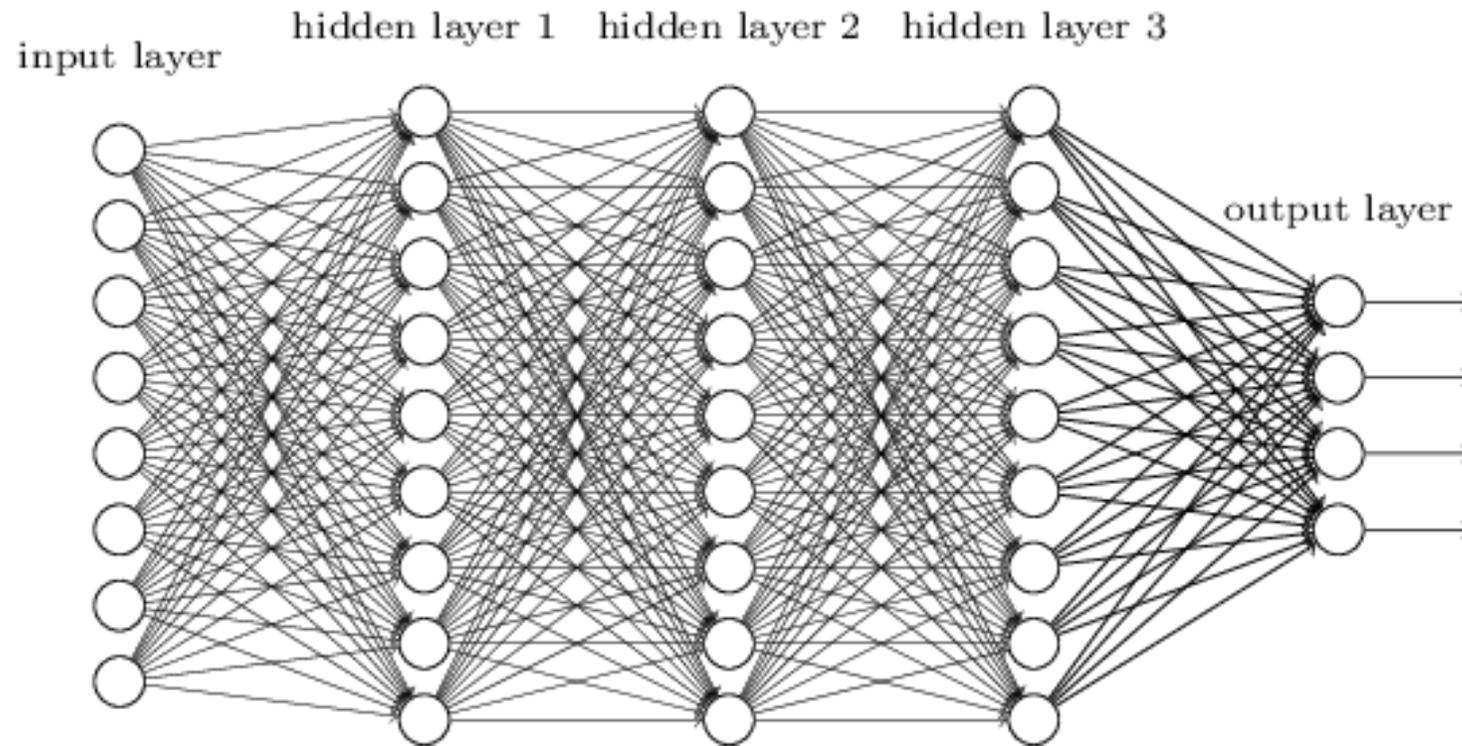
# Neurone Artificiale

Modello matematico (semplificato) del neurone biologico

$$y = f\left(\sum_{i=1}^n w_i x_i - \theta\right)$$

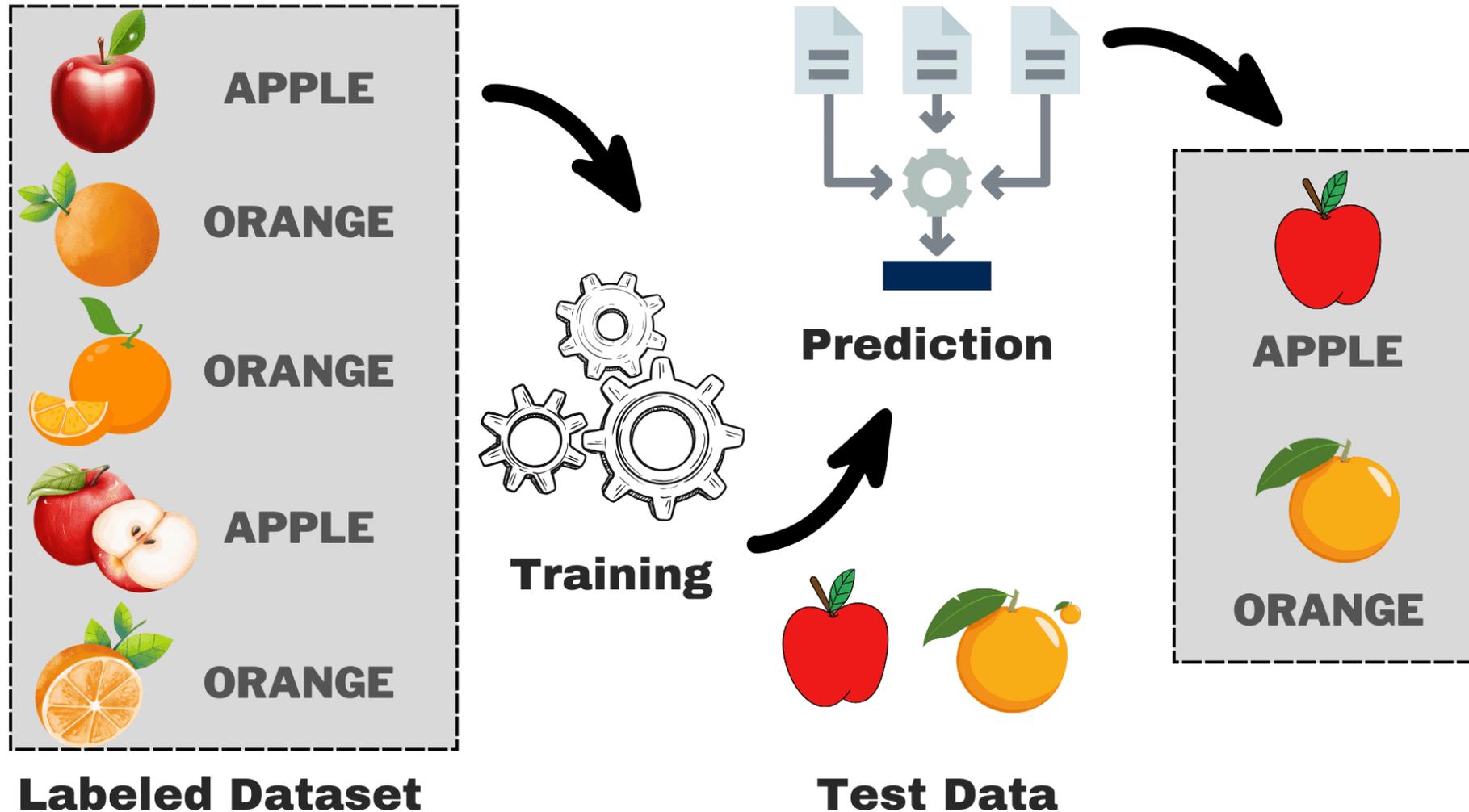
- Il neurone esegue la somma pesata degli  $n$  input ( $x_i$ , con pesi  $w_i$ ), che costituisce l'analogo del «potenziale di attivazione» del neurone;
- la funzione non lineare  $f$  (funzione di attivazione) con ingresso a soglia ( $\theta$ ) trasforma la somma pesata degli ingressi nel valore di uscita ( $y$ );

# Multi-Layer Perceptron (MLP)

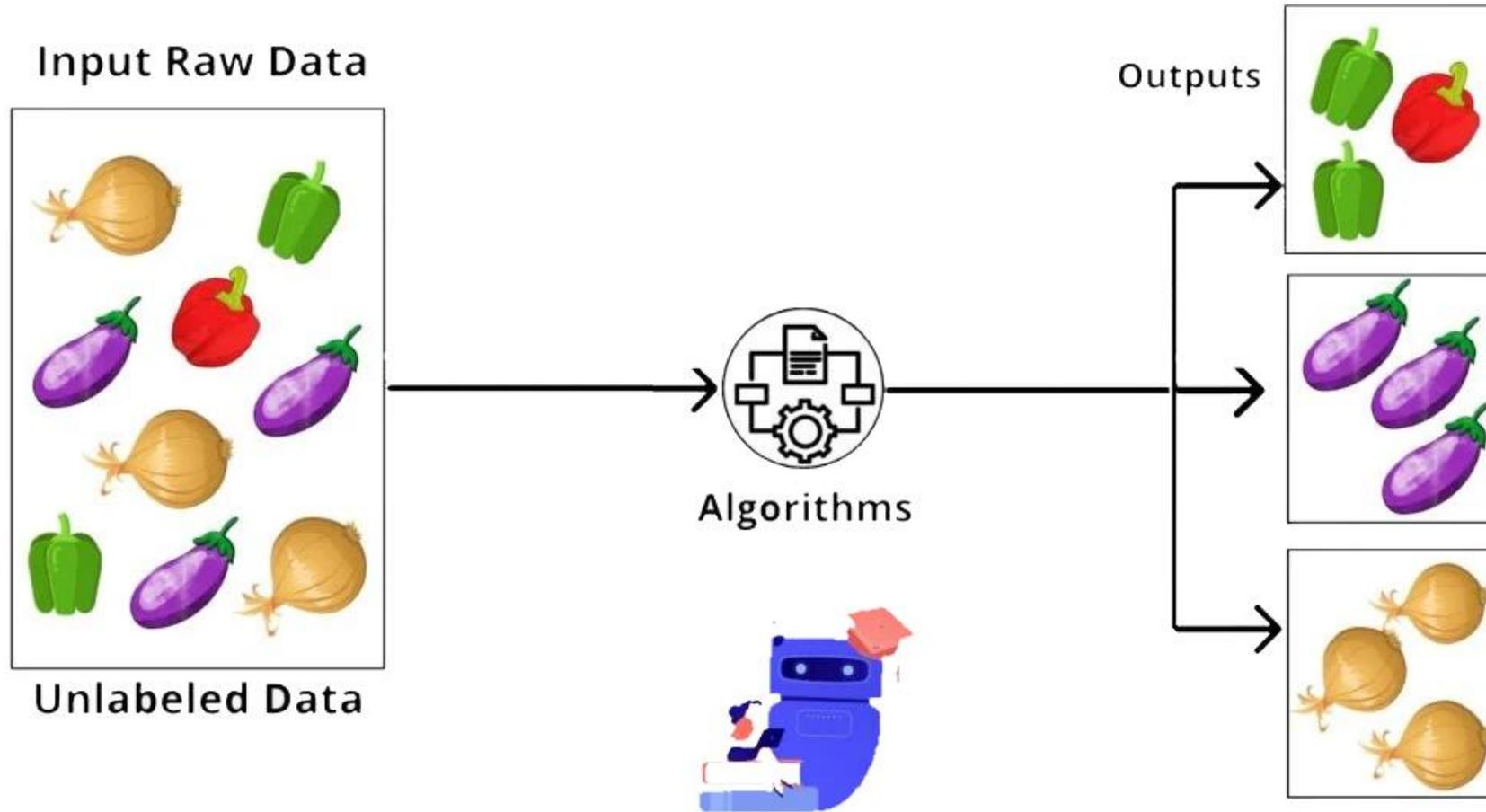


Da <http://neuralnetworksanddeeplearning.com/chap5.html>

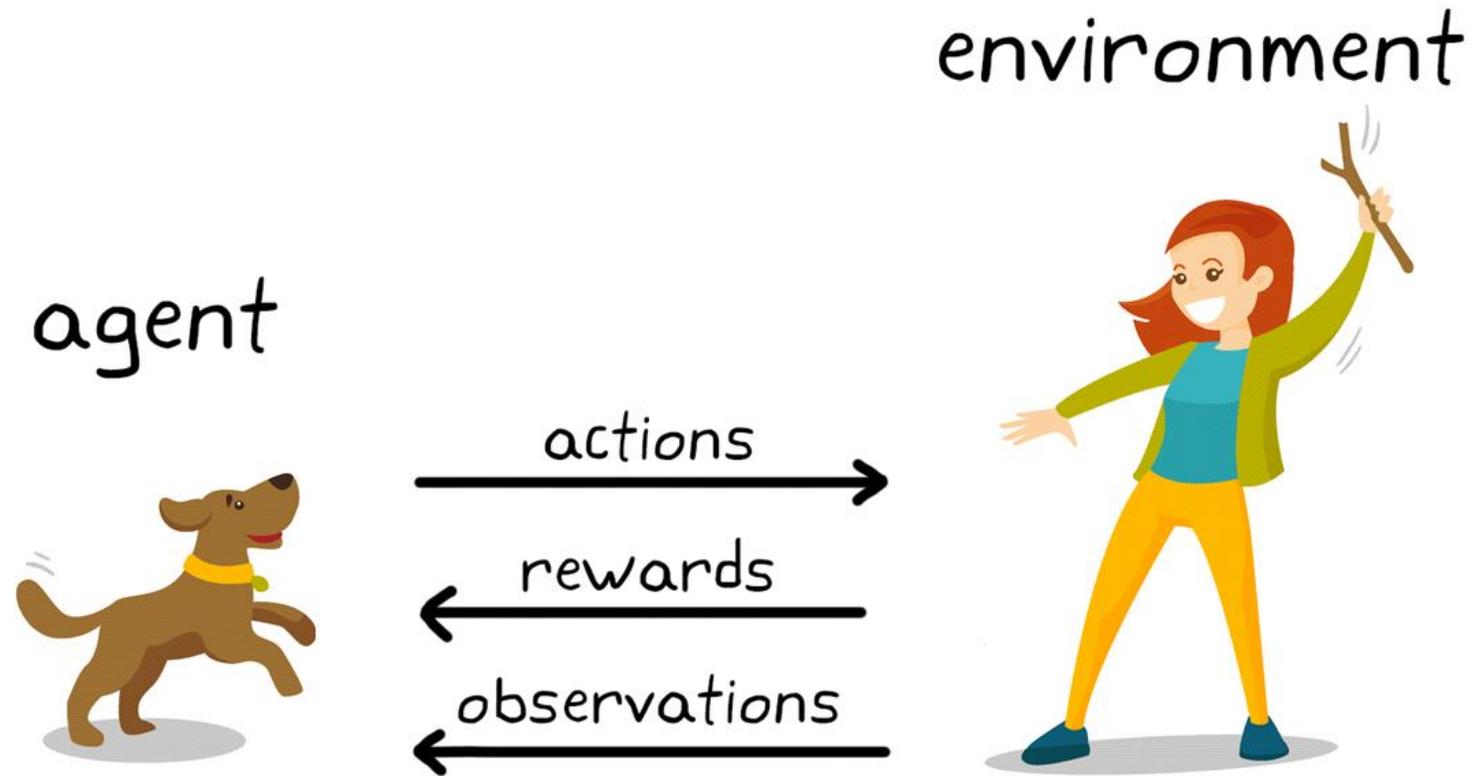
# Apprendimento supervisionato



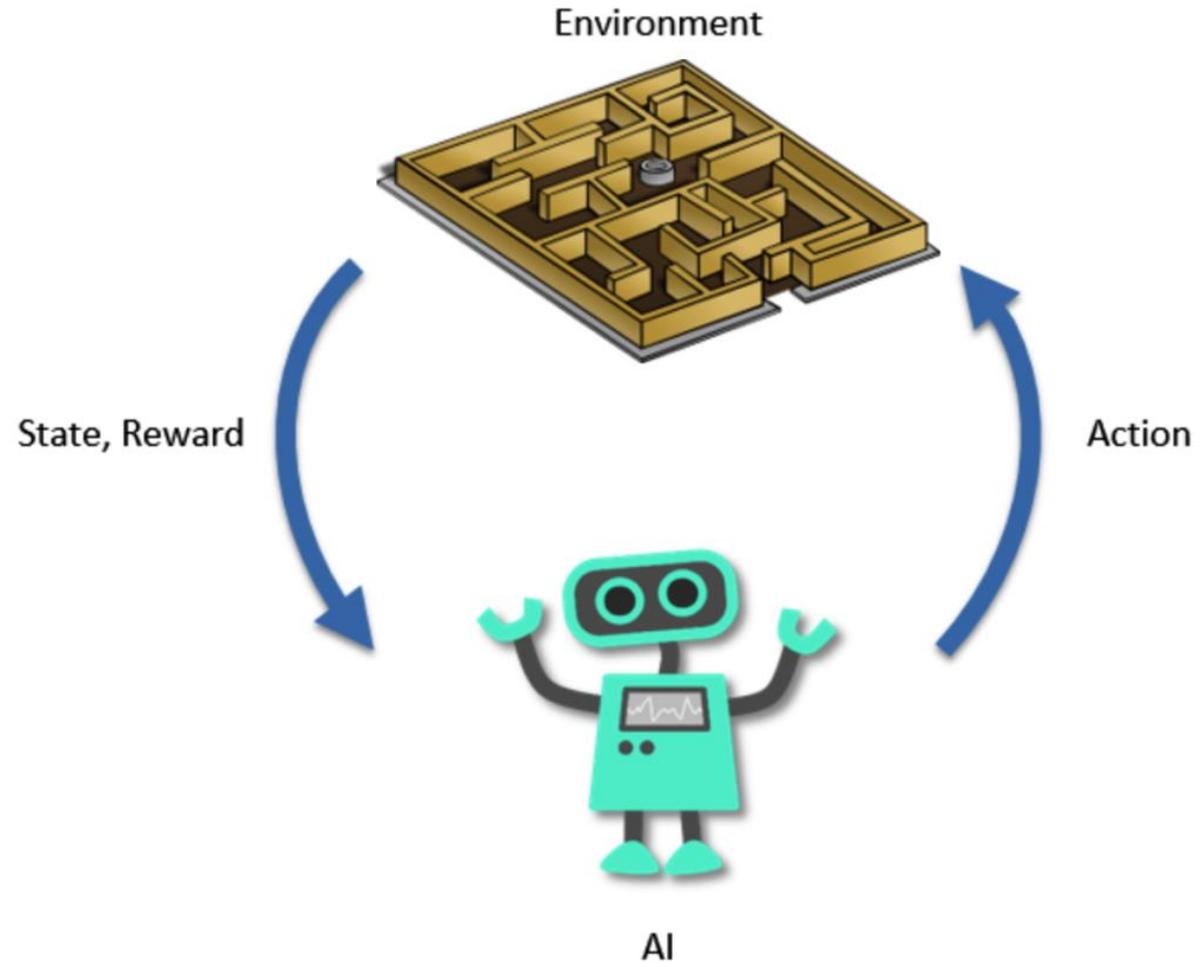
# Apprendimento non supervisionato



# Apprendimento basato su rinforzo



# Apprendimento basato su rinforzo



<https://www.youtube.com/watch?v=kopoLzvh5jY>

# Apprendimento supervisionato

Alla rete viene sottoposto un insieme di addestramento (training set) composto da coppie  $\langle \mathbf{x}, t \rangle$  dove  $\mathbf{x}$  è un ingresso (un insieme di valori, es. i pixel di un'immagine) e  $t$  è l'uscita desiderata

- La rete calcola l'output a partire dall'ingresso  $\mathbf{x}$ ;
- Sulla base della differenza tra l'output desiderato  $t$  e l'output effettivo della rete, la rete modifica i propri pesi per minimizzare tale differenza
- Il training set è sottoposto più volte alla rete (epoche di apprendimento) fintanto che tutti gli esempi sono classificati correttamente o l'errore scende sotto una certa soglia

# Error Back Propagation (EBP)

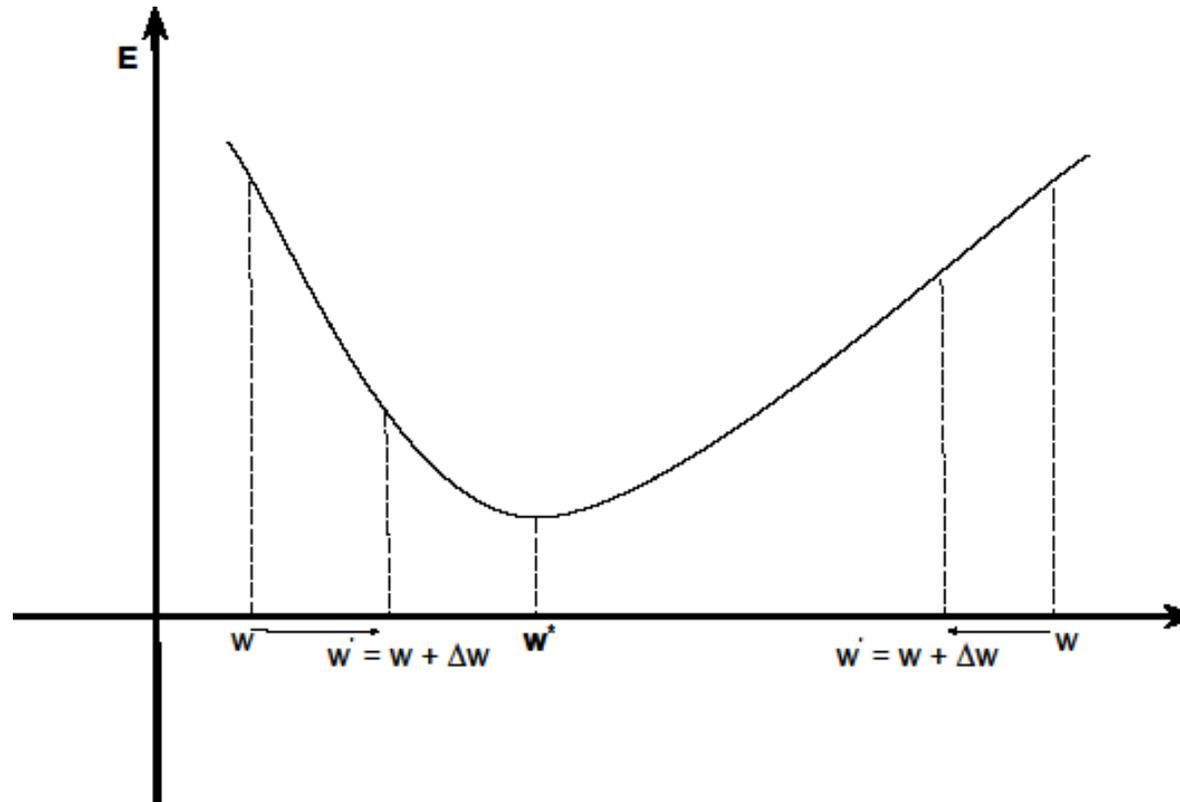
$$E(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^L \left( t_j^{(i)} - y_j^{(i)} \right)^2$$

dove:

- $y_j^{(i)}$  è l'uscita calcolata dal  $j$ -esimo neurone di uscita a fronte dell' $i$ -esimo ingresso della rete (ed è quindi funzione dei pesi della rete);
- $N$  è il numero di coppie  $\langle \mathbf{x}, \mathbf{t} \rangle$  «input, uscita desiderata» che compongono il training set della rete;
- $L$  è il numero di neuroni che compongono lo strato di uscita della rete.

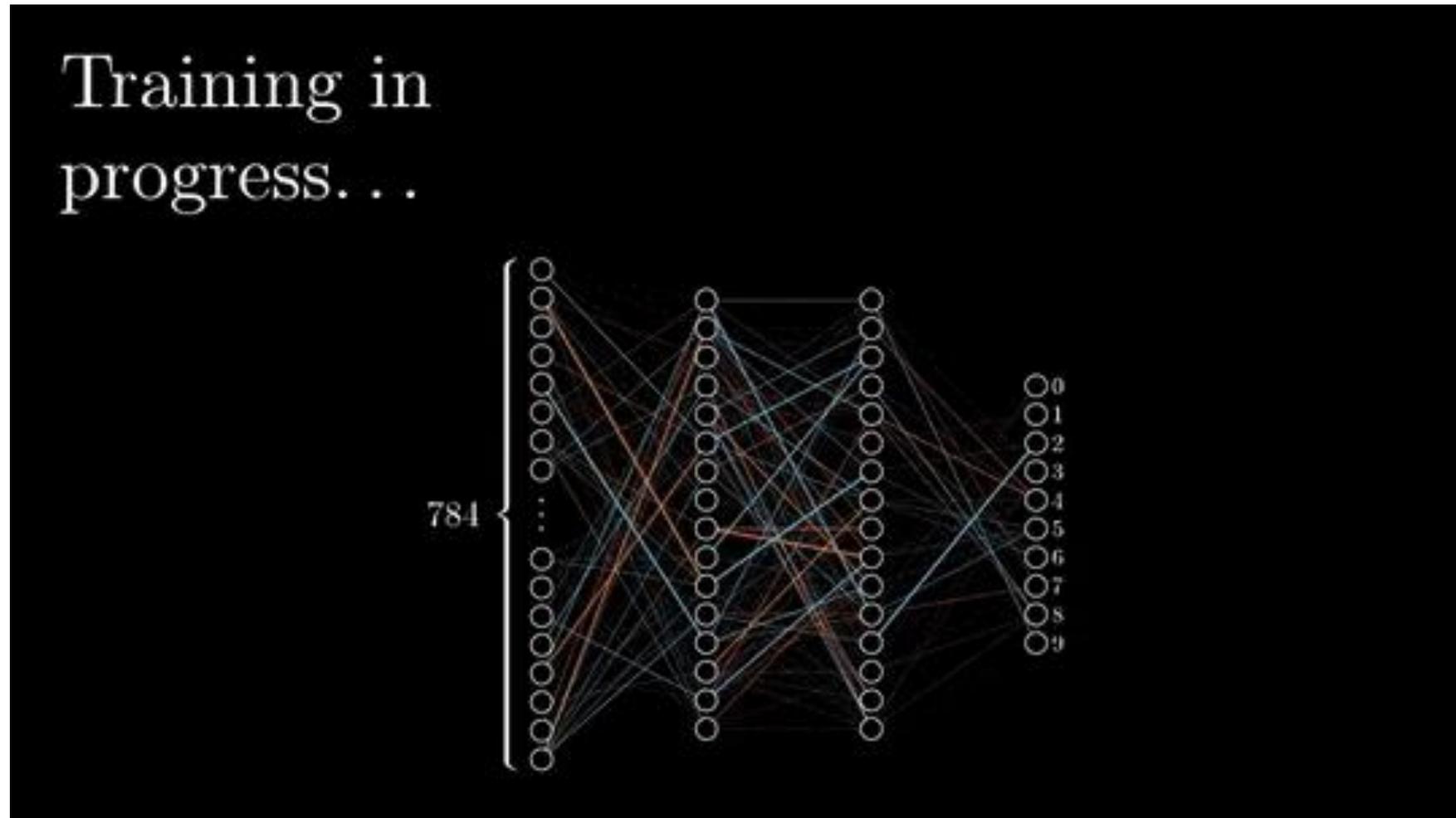
# Error Back Propagation (EBP)

L'addestramento consiste nella ricerca del valore dei pesi ( $w^*$ ) che minimizza la funzione di errore definita





# Error Back Propagation (EBP)



# Apprendimento - Esempio

Per l'addestramento di modello con apprendimento supervisionato, occorre preparare un insieme (un «dataset») composto da coppie *<dati di input, uscita desiderata>*.

Es. Riconoscimento cifre scritte a mano da immagini.

Per ogni immagine, «l'uscita desiderata» è il numero scritto all'interno



# Apprendimento - Esempio

Es. Database «MNIST» (Modified National Institute of Standards and Technology handwritten digit database)

- Immagini 28x28 in scala di grigi
- 60.000 immagini di training
- 10.000 immagini di test

Le immagini di validazione possono essere rimuovendone alcune dal training e riservandole per la validazione (per esempio riservandone a questo scopo un 10% dell'insieme di training, cioè usandone 54.000 per l'addestramento e 6.000 per la validazione).

## Apprendimento - Esempio

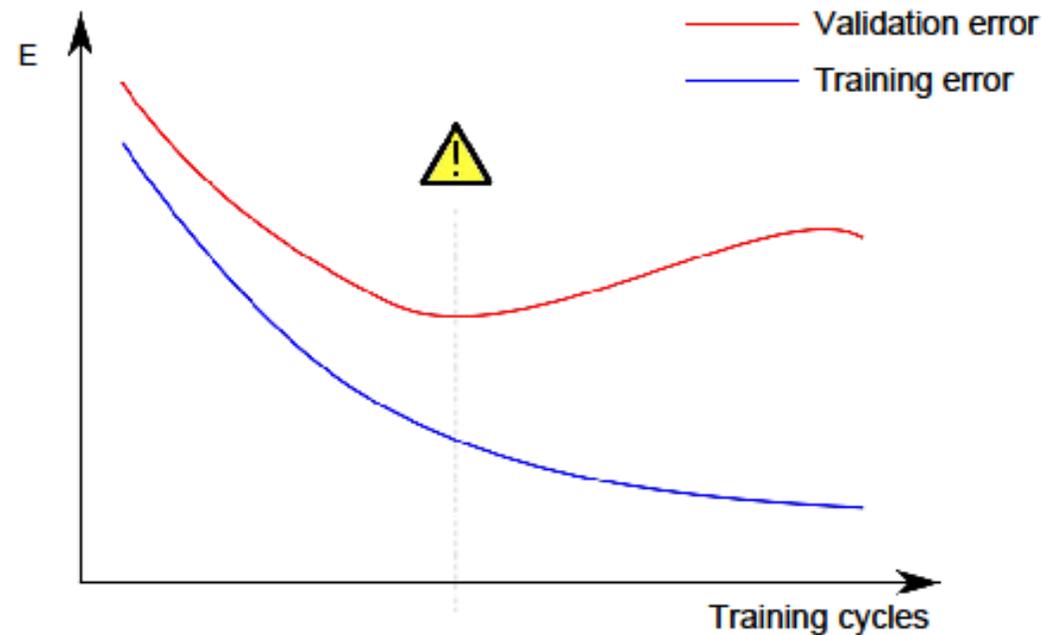
Quanti dati servono perché l'apprendimento sia efficace?

Non c'è una risposta diretta a questa domanda. Molti problemi di riconoscimento di immagini richiedono centinaia di migliaia o milioni di esempi di addestramento.

Dipende da molti fattori: complessità del problema, numero di feature (cioè cardinalità di ogni esempio dato in input), parametri del modello...

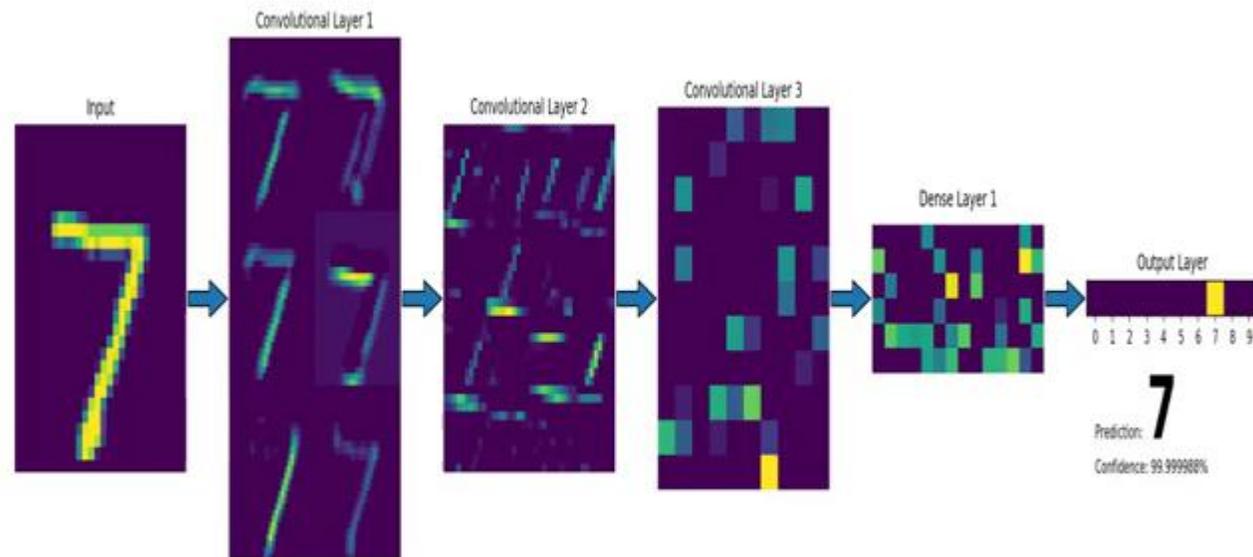
# Addestramento di un rete neurale

Se il numero di epoche di addestramento è troppo grande (e/o forniamo dati ridondanti, troppo simili tra loro), si può incorrere nell'overfitting



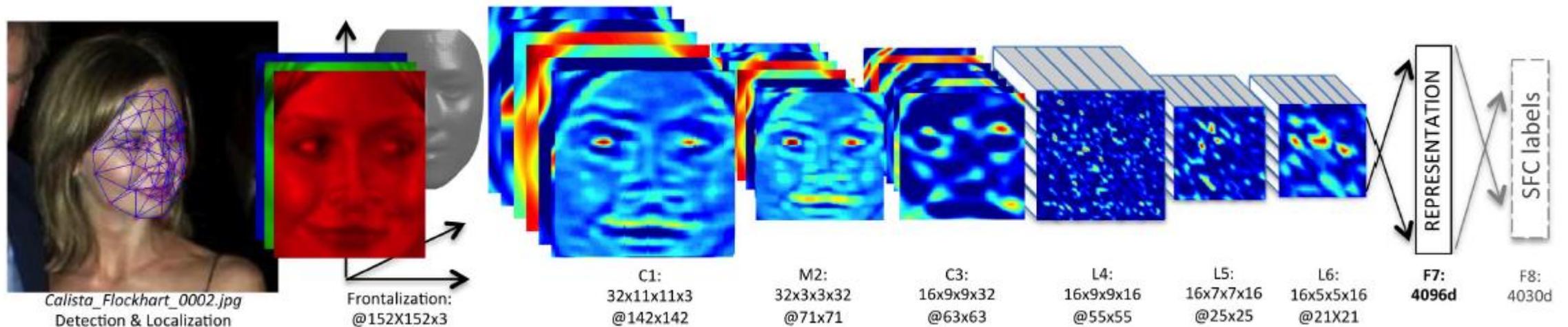
# Deep Learning e Deep Neural Network

Il termine “deep” in Deep Learning e Deep Neural Network, si riferisce alla presenza di molteplici strati in cascata ottenendo rappresentazione dei dati di input a livelli sempre più astratti.



# Esempio: Facebook DeepFace

Rete neurale per il riconoscimento facciale



Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification" in 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1701|1708

# Test di un rete neurale

Quali metriche usare per valutare le performance?

Accuratezza: numero di campioni classificati correttamente su numero di campioni totali

Es. Se il training set è composto da 100 cifre scritte a mano e se ne sono identificate correttamente 75 allora l'accuratezza è:

$$\frac{75}{100} = 0,75 = 75\%$$

# Test di un rete neurale

In un problema di classificazione binario, cioè costituito da due sole classi (positiva e negativa), possiamo indicare con:

- $T_p$  i veri positivi, cioè i campioni della classe positiva identificati correttamente
- $T_n$  i veri negativi, cioè i campioni della classe negativa identificati correttamente
- $F_p$  i falsi positivi, cioè i campioni della classe negativa erroneamente identificati come positivi
- $F_n$  i falsi negativi, cioè i campioni della classe positiva erroneamente identificati come negativi

# Test di un rete neurale

Precisione (precision): la porzione di positivi identificati correttamente tra tutti i positivi identificati

$$\frac{T_p}{T_p + F_p}$$

Recupero (recall): la porzione di positivi identificati correttamente tra tutti i positivi presenti nel test set

$$\frac{T_p}{T_p + F_n}$$

La precisione è una misura di «esattezza», il recupero una di «completezza»

# Test di un rete neurale

Supponiamo che un dataset sia composto da 50 esempi positivi e 50 esempi negativi e che la rete identifichi correttamente 47 esempi positivi e 48 esempi negativi. Questo significa che:

- Ci sono 47 veri positivi ( $T_p = 47$ )
- Ci sono 48 veri negativi ( $T_n = 48$ )
- Ci sono 3 falsi negativi ( $F_n = 3$ )
- Ci sono 2 falsi positivi ( $F_p = 2$ )

L'accuratezza è  $\frac{47+48}{100} = 0,95 = 95\%$

## Test di un rete neurale

- La precisione è  $\frac{47}{47+2} = 0,96 = 96\%$
- Il recupero è  $\frac{47}{47+3} = 0,94 = 94\%$

Supponiamo di avere un altro dataset di 100 campioni, 30 positivi e 70 negativi. Con la seguente classificazione:

$$T_p = 20, T_n = 50, F_p = 20, F_n = 10$$

Avremmo l'70% di accuratezza, 50% di precisione, 67% di recall.

**Con un dataset non bilanciato, cioè con un numero molto diverso di esempi per ogni classe, l'accuratezza non è una buona metrica!**

# Test di un rete neurale

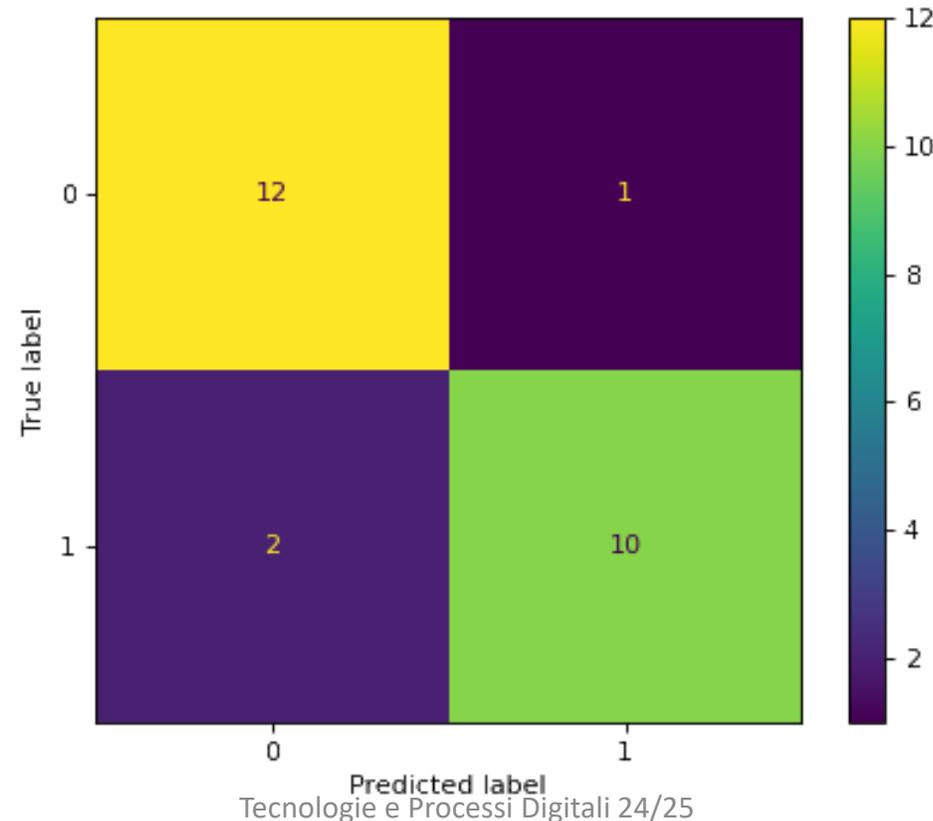
F1-score: media armonica di precisione e recupero

$$2 \frac{\textit{precisione} \times \textit{recupero}}{\textit{precisione} + \textit{recupero}}$$

Nell'esempio della slide precedente l'F1-score è  $0,57 = 57\%$

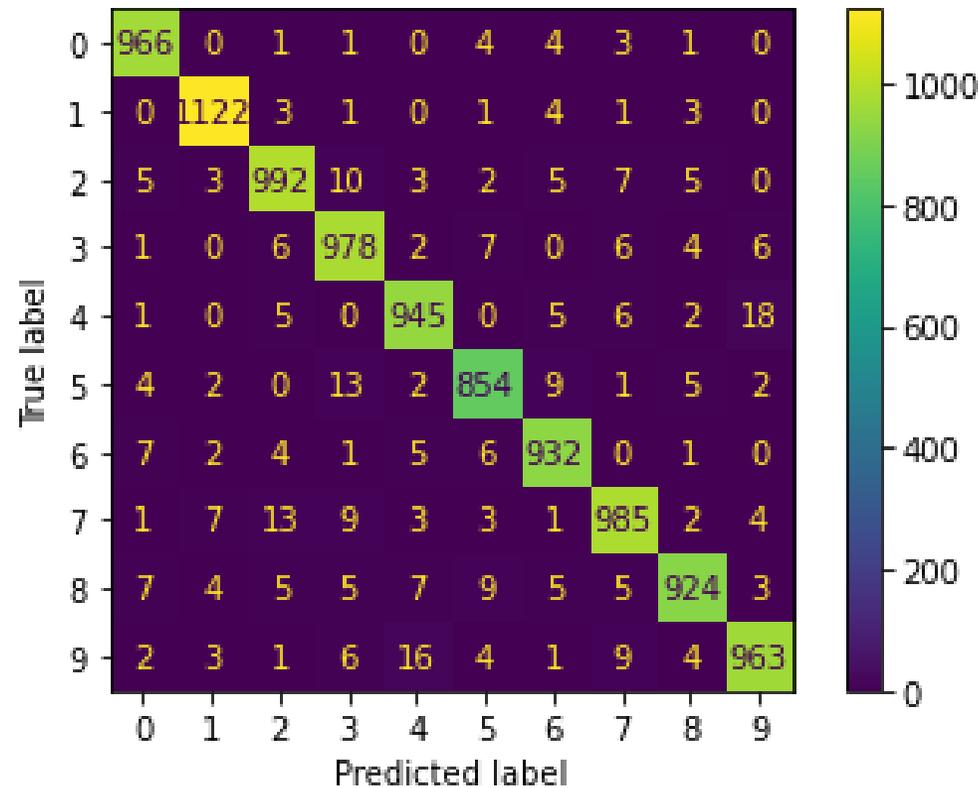
# Test di un rete neurale

Matrice di confusione: visualizzazione dell'accuratezza (e degli errori) di un classificatore. Sulle ordinate ci sono le classi reali, sulle ascisse le classi identificate dal classificatore.



# Test di un rete neurale

Es.: Identificazione delle cifre scritte a mano. Sulle ordinate c'è la cifra «reale», sulle ascisse quella identificata da una rete neurale



# Test di un rete neurale

Spesso, il training set e il test set vengono ottenuti prendendo dati dallo stesso dataset. I risultati ottenuti potrebbero dipendere dalla particolare divisione del dataset effettuata.

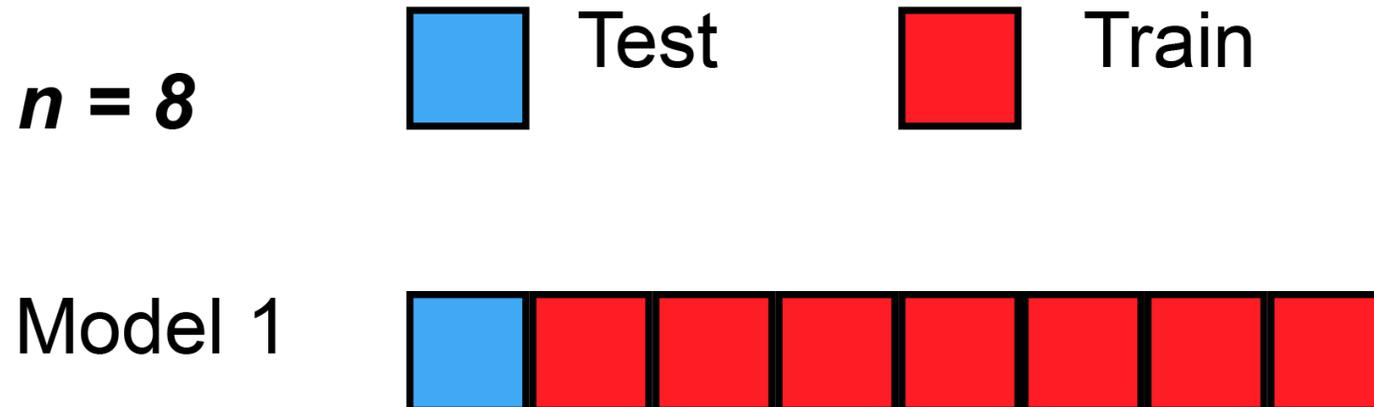
Per ovviare a questo problema si può ricorrere alla convalida incrociata

Convalida incrociata: usare differenti porzioni dei dati per il training e il test in differenti iterazioni di addestramento e valutazione, mediando poi i risultati finali.

# Test di un rete neurale

Una delle tecniche di convalida incrociata più usate è la *k*-fold Cross Validation:

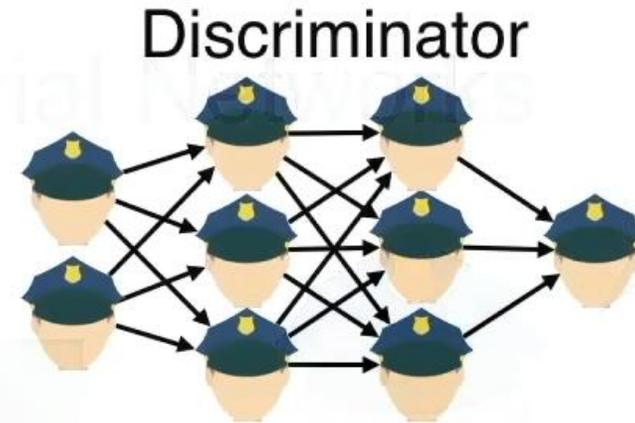
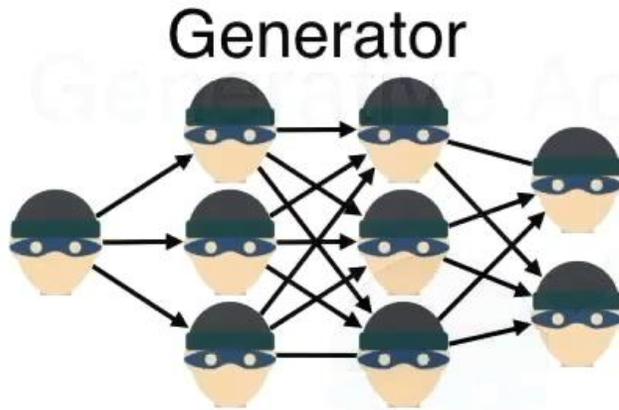
- Si divide il dataset in  $k$  sottoinsiemi di uguale dimensione
- Si eseguono training e test per  $k$  iterazioni, cioè  $k$  volte. Ad ogni iterazione si sceglie uno dei  $k$  sottoinsiemi per il test, e i restanti  $k-1$  per il training. All'iterazione successiva si cambia la scelta.





# Intelligenza Artificiale Generativa

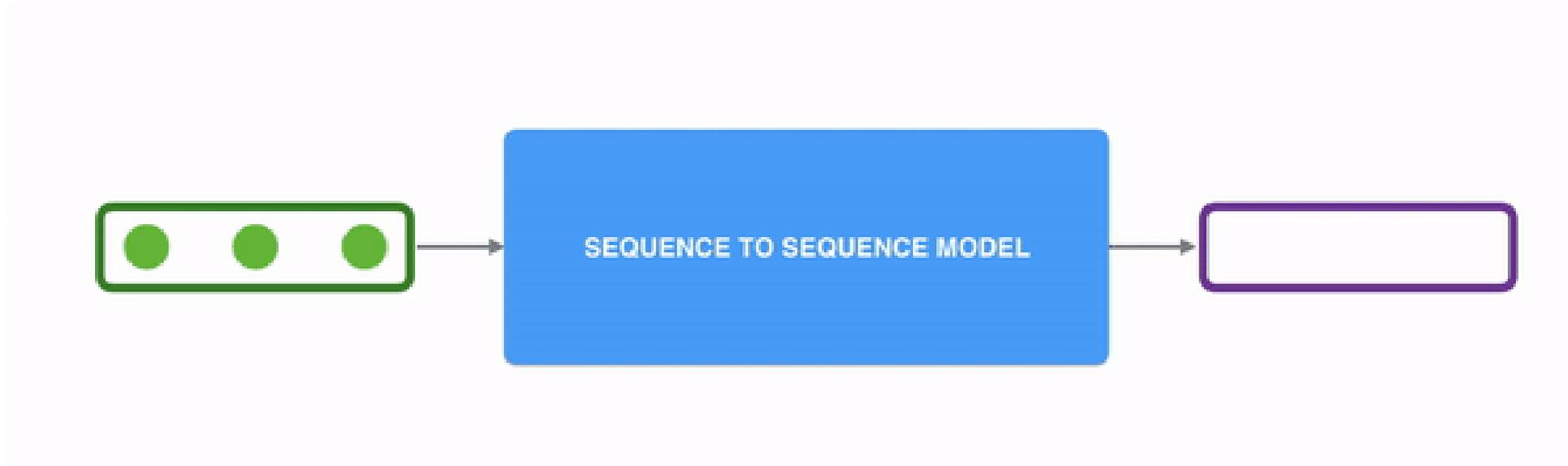
# GAN



Real images



# Transformers



# Generative Pretrained Transformer - GPT



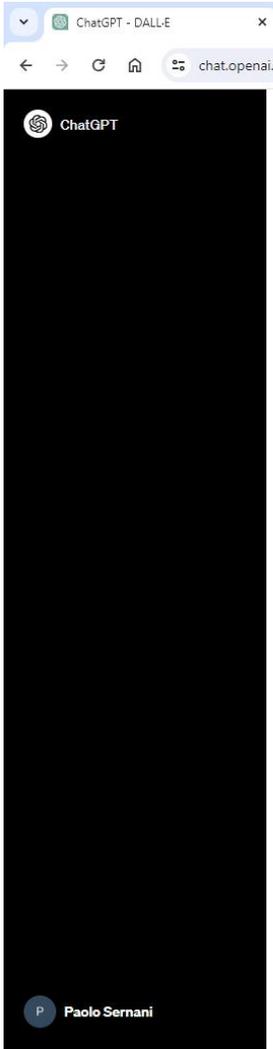
- Text generator basato su Deep Learning addestrato da OpenAI
- Basato su Transformer
- Next word prediction su 300 miliardi di token
- 175 miliardi di parametri (800 GB)

# Creazione immagini





# Creazione immagini



You

In un cartone Disney Pixar, una ragazza con i capelli lunghi e gli occhiali, mentre studia un libro sulla sua scrivania, scorge dalla finestra suo nonno, avvolto in un cappotto marrone e una sciarpa rossa, mentre cammina sulla neve, e lo saluta



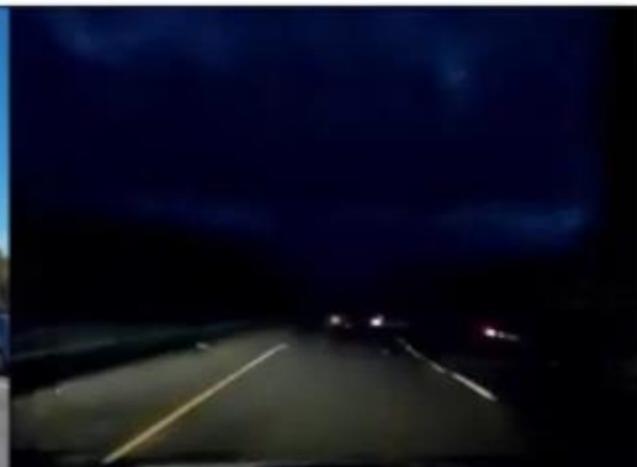
DALL-E



Here are the images depicting the scene you described, with the girl waving at her grandfather from the window.









# ChatGPT

Chatbot basato su GPT (3.5, 4)

- Comprensione del linguaggio umano
- Risposte testuali che imitano quelle di un umano
- Risposte «accurate»: contestualmente appropriate alla richiesta ricevuta



# Che cos'è ChatGPT?



*A chatbot is an AI system and one of the most elementary and widespread examples of intelligent Human-Computer Interaction (HCI): a computer program, which responds like a smart entity when conversed with through text or voice and understands one or more human languages by Natural Language Processing (NLP). It is defined as “A computer program designed to simulate conversation with human users”*

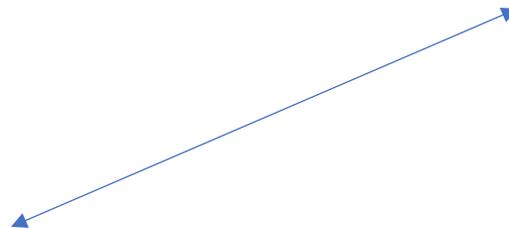
Da Adamopoulou, E., & Moussiades, L. (2020). An overview of chatbot technology. In *IFIP international conference on artificial intelligence applications and innovations* (pp. 373-383). Springer, Cham.  
[https://link.springer.com/chapter/10.1007/978-3-030-49186-4\\_31?ref=blog.min.io](https://link.springer.com/chapter/10.1007/978-3-030-49186-4_31?ref=blog.min.io)

# Modellare il linguaggio

Che cosa segue la frase «Ho cenato, ho visto un film e ho preparato...»



- Una tisana
- Un panino
- Il letto
- I vestiti per domani
- La lista delle cose da fare
- Un bicchiere di latte
- L'outfit
- ...



# Modellare il linguaggio

Che cosa segue la frase «Ho cenato, ho visto un film e ho preparato...»



Le parole che seguono la frase dipendono:

- Dalle abitudini di chi le pronuncia
- Dal suo mood
- Dal contesto

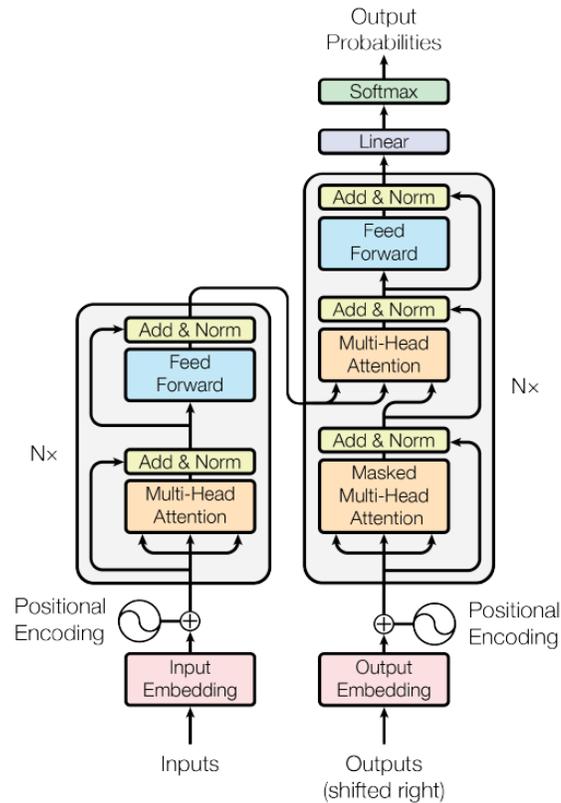
# Word embeddings

Il linguaggio naturale e il suo «processamento» (Natural Language Processing – NLP) richiedono una rappresentazione numerica



Word embeddings = rappresentazione numerica delle parole in un testo

# Transformers



- Scala (GPU multi-core)
- Elabora in parallelo i dati di input
- Meccanismo di «attenzione»

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.

[https://proceedings.neurips.cc/paper\\_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html](https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html)

# Generative Pretrained Transformer - GPT



- Text generator basato su Deep Learning addestrato da OpenAI
- Basato su Transformer
- Next word prediction su 400 miliardi di token
- 175 miliardi di parametri (800 GB)

# ChatGPT - Allucinazioni

## Frase «plausibili», ma irrilevanti o errate fattualmente

PA summarise this article <https://www.nytimes.com/2023/03/11/technology/chatgpt-prompts-to-avoid-content-filters.html>



The article discusses how ChatGPT, an AI language model, is being used to circumvent content filters that are designed to detect and block certain types of content. By generating prompts that are less likely to trigger these filters, ChatGPT can be used to create content that might otherwise be censored. While this technology could be useful for those seeking to express themselves freely, it also raises concerns about the potential misuse of AI-generated content, as well as the difficulty of regulating such content in the future. The article concludes by noting that there is still much to learn about the impact of AI language models like ChatGPT on the way we communicate online.

Immagine da: [https://en.wikipedia.org/wiki/Hallucination\\_\(artificial\\_intelligence\)](https://en.wikipedia.org/wiki/Hallucination_(artificial_intelligence))

# ChatGPT - Limiti

- Non sempre dice il vero!
- Limite di parole
- Serve il contesto!



# Prompt engineering

- Prompt: istruzioni e contesto passati a un LLM per ottenere l'esecuzione del compito desiderato
- Prompt engineering: lo sviluppo e l'ottimizzazione dei prompt per usare «efficacemente» un LLM

Istruzioni e contesto

Input

Output

# Contesto

- Dettagli rilevanti rispetto alla richiesta aiutano ad ottenere risposte accurate



Esempio: vendita di un  
prodotto

# Contesto

Per ottenere informazioni specifiche, fornire del contenuto aiuta!

- Conoscenza di dominio: includere contenuto per permettere a ChatGPT di usare una terminologia specifica
- Consistenza del tono
- Linee guida e compliance



Esempio: social media

# Prompt engineering

- Dire chiaramente l'obiettivo/task (es. «Scrivi delle strategie di marketing» vs «Proponi tre diverse strategie di marketing per piccole imprese»)
- Settare parametri: quante parole, caratteri frasi, quante alternative, quale punto di vista, quale formato
- «Stepwise refinement»!



Esempi

# Prompt engineering - Sintesi

1. Nuova conversazione
2. Obiettivo specifico (chiedere di assumere il ruolo di un qualche tipo di esperto)
3. Richiedere un compito il più specifico possibile
4. Condizioni che ChatGPT deve rispettare
5. (Chiedere riguardo la comprensione, per chiarire eventuali punti ambigui)

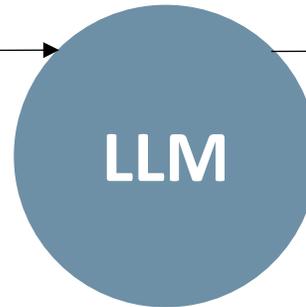


# Prompt

Prompt

Da che cosa è composto un piatto di carbonara?

Contesto



Completion

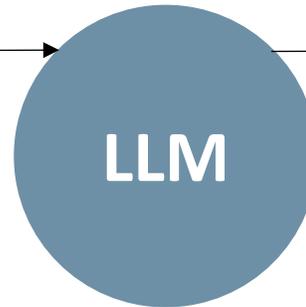
Da che cosa è composto un piatto di carbonara?

La carbonara è un piatto classico della cucina italiana, in particolare di quella romana. Gli ingredienti principali di una tradizionale carbonara sono pasta, guanciale, uova, pecorino romano, pepe nero, sale.

# Apprendimento basato sul contesto

Prompt

Classifica questo commento:  
Il libro è stupendo!  
Sentiment:



Completion

Classifica questo commento:  
Il libro è stupendo!  
Sentiment: Positivo

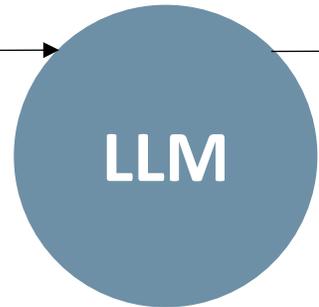
Zero-shot inference

# Apprendimento basato sul contesto

Prompt

Classifica questo commento:  
Il libro è stupendo!  
Sentiment: **Positivo**

Classifica questo commento:  
Non mi è piaciuto quel  
ristorante  
Sentiment:



Completion

Classifica questo commento:  
Il libro è stupendo!  
Sentiment: Positivo

Classifica questo commento:  
Non mi è piaciuto quel  
ristorante  
Sentiment: **Negativo**

One-shot inference

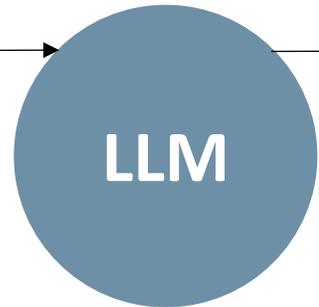
# Apprendimento basato sul contesto

## Prompt

Classifica questo commento:  
Il libro è stupendo!  
Sentiment: **Positivo**

Classifica questo commento:  
Non mi è piaciuto quel ristorante  
Sentiment: **Negativo**

Classifica questo commento:  
La pizza era pessima  
Sentiment:



## Completion

Classifica questo commento:  
Il libro è stupendo!  
Sentiment: Positivo

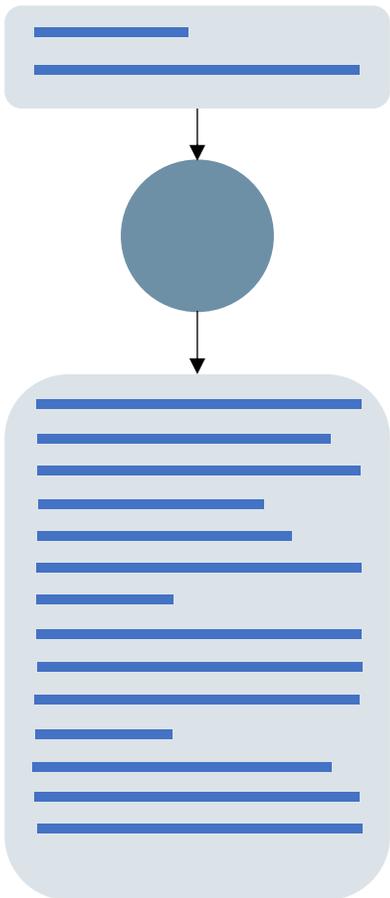
Classifica questo commento:  
Non mi è piaciuto quel ristorante  
Sentiment: Negativo

Classifica questo commento:  
La pizza era pessima  
Sentiment: **Negativo**

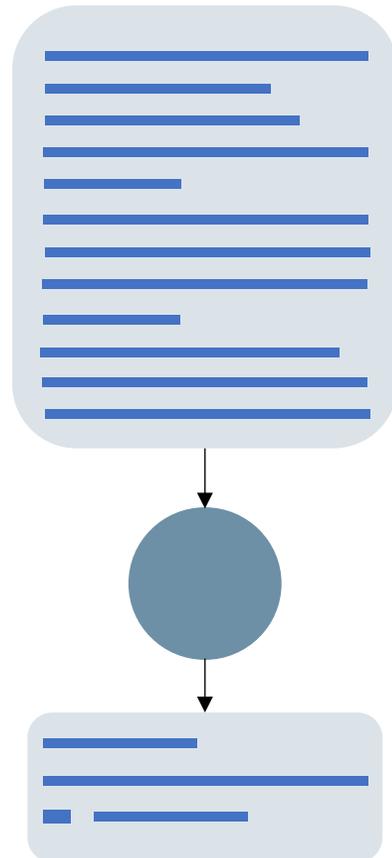
## Few-shots inference

# LLM – Casi d'uso

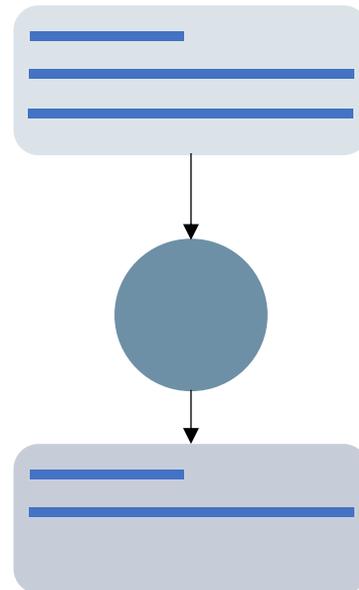
Generazione testo



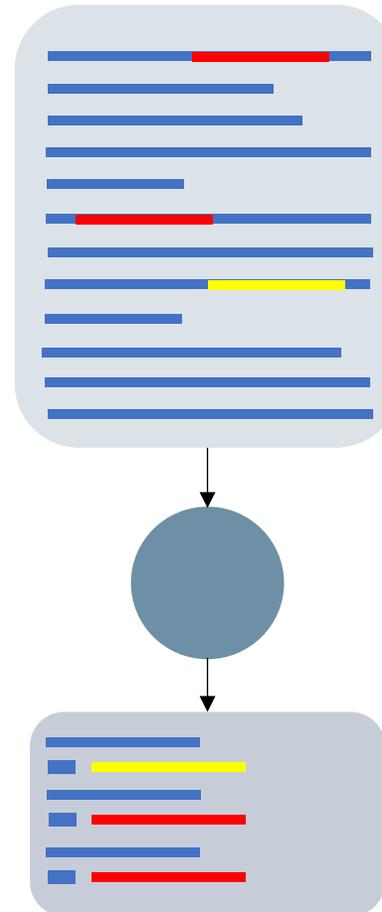
Sintesi e parafrasi



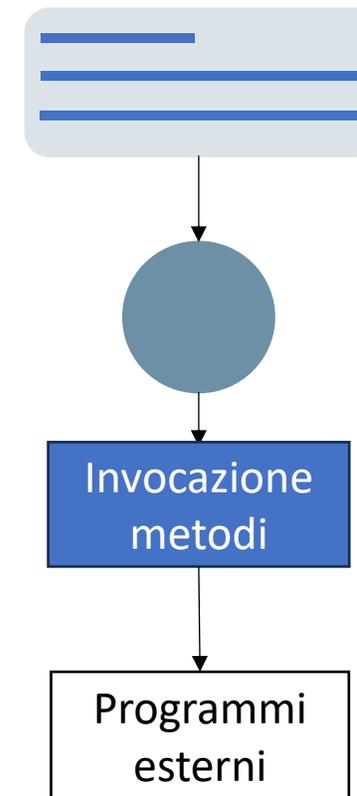
Traduzione



Information retrieval



Programmazione e APIs



## (Alcune) Criticità

# Annotazione dei dataset (1)

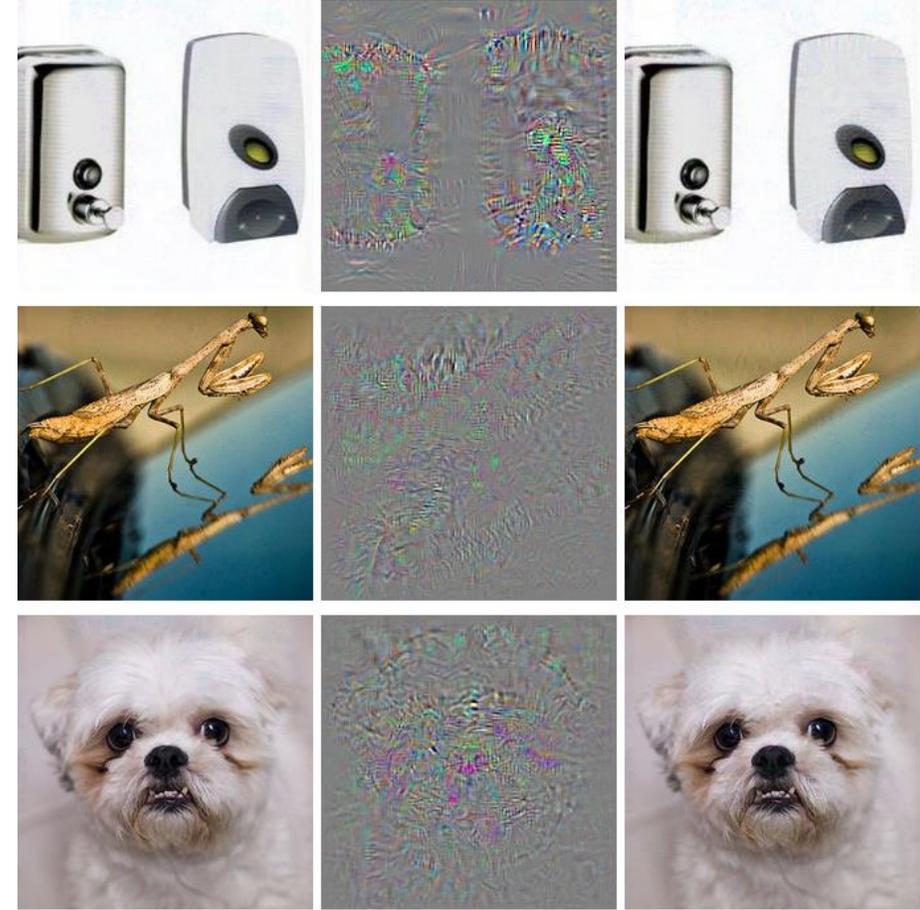
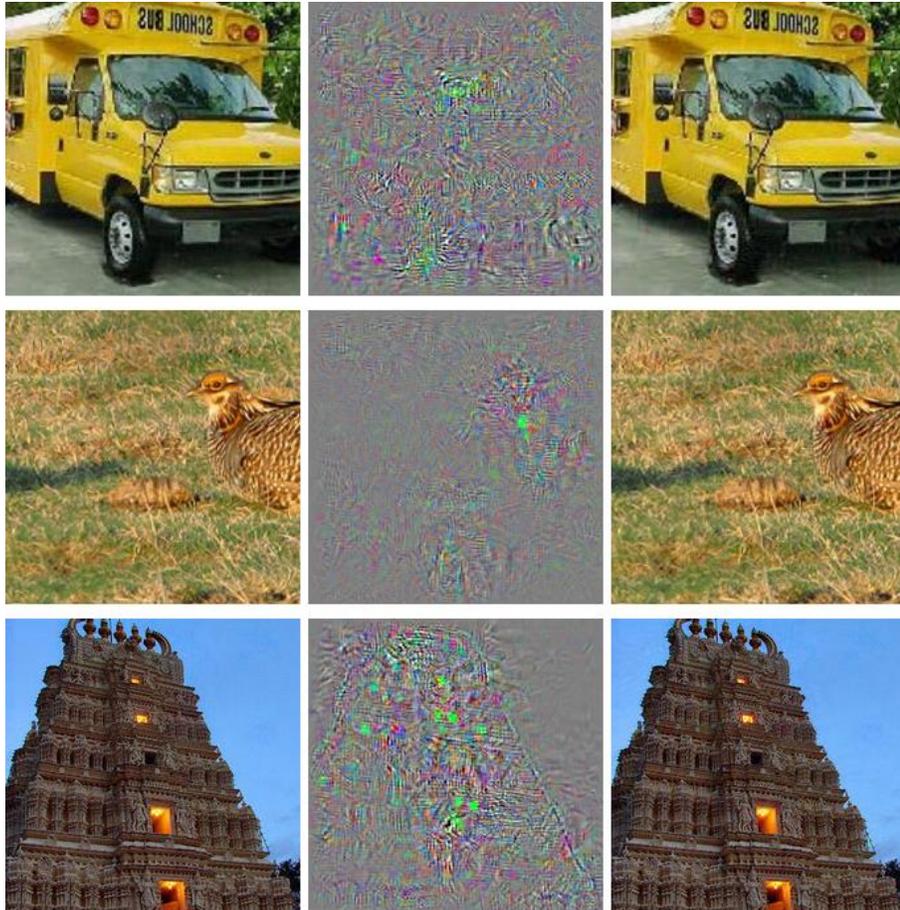


SO MUCH OF "AI" IS JUST FIGURING OUT WAYS TO OFFLOAD WORK ONTO RANDOM STRANGERS.

# Annotazione dei dataset (2)

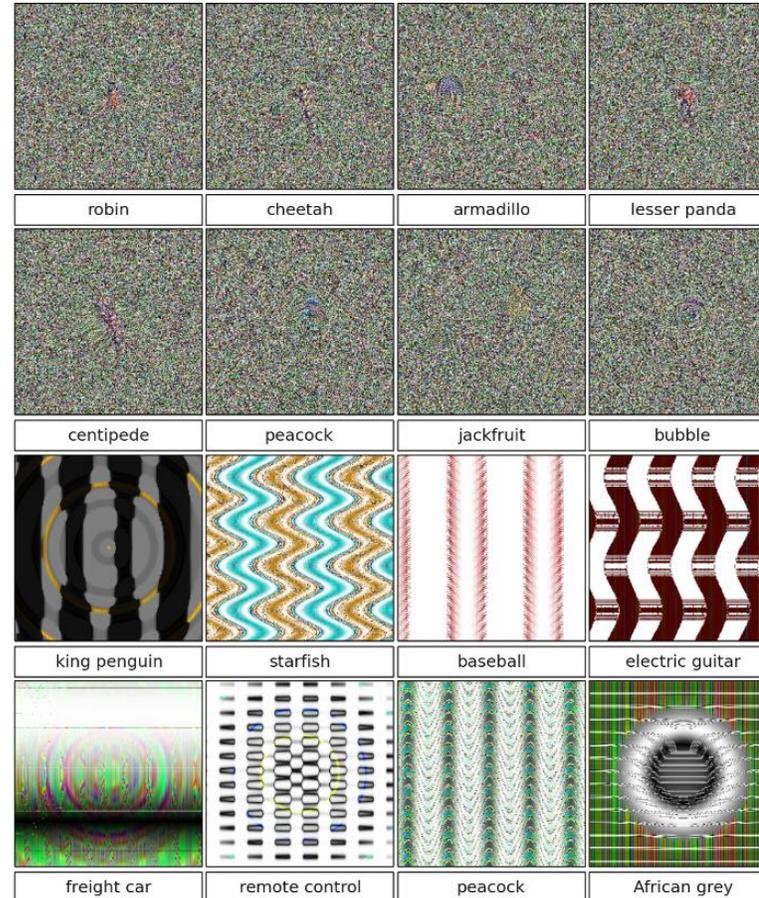
The screenshot displays two side-by-side annotation tasks. The left task, titled "Select all images with commercial lorries", shows a grid of nine images. Three images are marked with a blue checkmark: a red truck, a white truck with "ISOLATION PLUS" on its side, and a white truck with a red stripe. The right task, titled "Select all squares with fire hydrants", shows a 4x4 grid of 16 squares. The top-left square contains a red fire hydrant, while the other 15 squares are empty. Both tasks have a blue button at the bottom: "Verify" for the lorry task and "SKIP" for the fire hydrant task. Navigation icons (refresh, headphones, info) and a "Report a Problem" link are visible at the bottom of each task.

# Potenziali problemi (1)



C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, *Intriguing properties of neural networks*, arXiv preprint arXiv:1312.6199, 2013. <https://arxiv.org/abs/1312.6199>

# Potenziali problemi (2)



A. Nguyen, J. Yosinski and J. Clune, *Deep neural networks are easily fooled: High confidence predictions for unrecognizable images*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 427-436.

# Deep Neural Networks

Grande accuratezza in molti compiti basati sul riconoscimento di immagini (e non solo...)

L'addestramento richiede molte risorse:

- Hardware
- Migliaia/Milioni di esempi di addestramento

Usate spesso come «scatole nere»: scarsa trasparenza nella decisione

## Black box - Esempio

Una rete neurale stabilisce che due immagini di volti corrispondono alla stessa persona. Questo perché una rete neurale da 140 milioni di parametri (pesi e soglie) addestrata ad estrarre dalle immagini vettori di 1024 valori compresi tra 0 e 1 che dovrebbero descrivere il volto, stabilisce che i due vettori che rappresentano le immagini sono «vicini».

Sarebbe sufficiente durante un dibattito per stabilire la persona in quelle due foto è la stessa?

# Reti neurali come black box

GDPR – **G**eneral **D**ata **P**rotection **R**egulation

Articolo 22 - Processo decisionale automatizzato relativo alle persone fisiche, compresa la profilazione\*

*«L'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona.»*

---

[\\*https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:32016R0679&from=IT#d1e2845-1-1](https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:32016R0679&from=IT#d1e2845-1-1)

# Trasparenza delle decisioni

La «trasparenza» delle tecniche di IA è coinvolta in alcuni articoli del GDPR (ad esempio nell'art. 13, oltre che nel 22)

Di fatto, occorrerebbe una spiegazione sul processo decisionale. Occorrerebbe cioè:

- capire come una certa tecnica o metodologia operano
- capire quali fattori sono stati decisivi per un certo output



eXplainable Artificial Intelligence – XAI

Per approfondire: <https://www.agendadigitale.eu/cultura-digitale/lintelligenza-artificiale-deve-essere-spiegabile-ecco-i-progetti-e-le-tecniche/>

## Bias nei dati

Elementi «sovra-rappresentati» nei dati possono portare output errati (inaspettati?)

Es.: algoritmi usati nel *job recruiting*

- <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>
- <https://www.forbes.com/sites/forbestechcouncil/2023/09/25/ai-bias-in-recruitment-ethical-implications-and-transparency/?sh=6a34e651799f>

## Microsoft Vulnerability Severity Classification for AI Systems

<https://www.microsoft.com/en-us/msrc/aibugbar>

Quanta energia per l'addestramento e l'uso di IA?

## Fattori chiave

- **Hardware:** tipo (ad esempio, quali GPU) ed efficienza
- **Dati di addestramento:** scala e complessità del dataset
- **Modello di IA/**Tipo di rete neurale
- Dove stanno i server e i data center?

## GPT 3 – 3.5

- Generatore di testo di OpenAI, basato su Transformer (175 miliardi di parametri ~ 800 GB in memoria)
- **Addestramenti:** next word prediction su 45 TB testo, 400 miliardi of tokens
- 1024 A100 GPU, 34 giorni

T. Brown et al., «Brown, Tom B. "Language models are few-shot learners." arXiv preprint arXiv:2005.14165 (2020). <https://arxiv.org/abs/2005.14165>

D. Narayanan et al. «Efficient large-scale language model training on gpu clusters using megatron-lm.» *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*. 2021.

# A100...



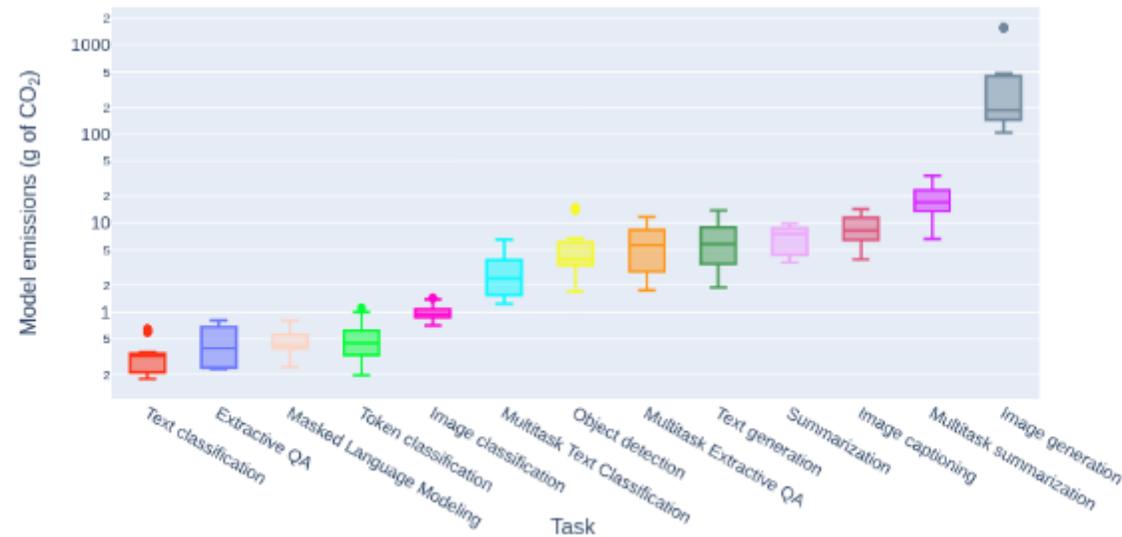
**NVIDIA A100  
TENSOR CORE GPU**

	NVIDIA A100 for NVIDIA HGX™	NVIDIA A100 for PCIe
GPU Architecture	NVIDIA Ampere	
Double-Precision Performance	FP64: 9.7 TFLOPS FP64 Tensor Core: 19.5 TFLOPS	
Single-Precision Performance	FP32: 19.5 TFLOPS Tensor Float 32 (TF32): 156 TFLOPS   312 TFLOPS*	
Max Power Consumption	400 W	250 W

## Power Hungry Processing: ⚡ Watts ⚡ Driving the Cost of AI Deployment?

ALEXANDRA SASHA LUCCIONI and YACINE JERNITE, Hugging Face, Canada/USA

EMMA STRUBELL, Carnegie Mellon University, Allen Institute for AI, USA



Luccioni, S., Jernite, Y., & Strubell, E. (2024, June). Power hungry processing: Watts driving the cost of AI deployment?. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 85-99). <https://arxiv.org/abs/2311.16863>

# Alcuni dati - Training

## **GPT 4**

- 25,000 A100 GPU, 100 giorni (28,800,000 kW/h)
- 6,912,000 CO2 t (1,300 homes per year)

## Llama 2

- 3,311,616 ore di GPU, (452 homes for a year)

## Alcuni dati – Inferenza

- 60% of energy usage for inference
- ChatGPT-4 query (0.001-0.01 kWh) – Google Search query (0.0003 kWh) 1567 % (78 billion queries per year, 3,500 homes per year)

D. Patterson *et al.*, "The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink," in *Computer*, vol. 55, no. 7, pp. 18-28, July 2022, doi: 10.1109/MC.2022.3148714.

<https://arxiv.org/pdf/2204.05149>

<https://medium.com/darrowai/code-green-addressing-the-environmental-impact-of-language-models-0161eb790c21>

# Responsible AI

- Scegliere il modello e l'architettura giusti: Gli LLM non sono adatti a tutto!
- Quantizzazione (riduzione dei modelli alle dimensioni necessarie nei dispositivi con emulazione)



Ricerca

# Human level AI?

# Artificial General Intelligence - AGI



Yann LeCun

17 maggio alle ore 15:20 · 🌐

About the raging debate regarding the significance of recent progress in AI, it may be useful to (re)state a few obvious facts:

- (0) there is no such thing as AGI. Reaching "Human Level AI" may be a useful goal, but even humans are specialized.
- (1) the research community is making \*some\* progress towards HLA
- (2) scaling up helps. It's necessary but not sufficient, because....
- (3) we are still missing some fundamental concepts
- (4) some of those new concepts are possibly "around the corner" (e.g. generalized self-supervised learning)
- (5) but we don't know how many such new concepts are needed. We just see the most obvious ones.
- (6) hence, we can't predict how long it's going to take to reach HLA.

I really don't think it's just a matter of scaling things up.

We still don't have a learning paradigm that allows machines to learn how the world works, like human and many non-human babies do.

Some may believe that scaling up a giant transformer trained on sequences of tokenized inputs is enough.

Others believe "reward is enough".

A few others believe that explicit symbol manipulation is necessary.

A few don't believe gradient-based learning is part of the solution.

I believe we need to find new concepts that would allow machines to:

- learn how the world works by observing, like babies.
- learn to predict how one can influence the world through taking actions.
- learn hierarchical representations that allows long-term predictions in abstract representation spaces.
- properly deal with the fact that the world is not completely predictable.
- enable agents to predict the effects of sequences of actions so as to be able to reason and plan
- enable machines to plan hierarchically, decomposing a complex task into subtasks.
- all of this in ways that are compatible with gradient-based learning.

The solution is not just around the corner.

We have a number of obstacles to clear, and we don't know how.