

Intelligenza artificiale: una tecnologia difficile da controllare

Alessandro Verri

MaLGA, DIBRIS, Università degli Studi di Genova

Intelligenza artificiale o machine learning?

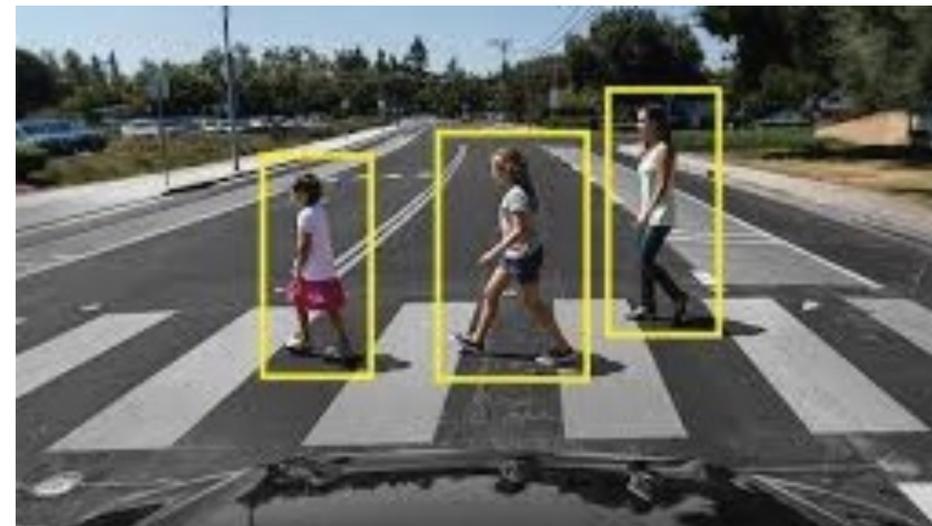
Da ormai vent'anni il **Machine Learning**, un'idea degli anni '60, consente lo sviluppo di tecnologie di enorme successo

Si addestra un computer a risolvere un problema per mezzo di esempi, o coppie specifiche di problema-soluzione, invece che determinare esplicitamente la sequenza di istruzioni necessarie

Esempi di esempi (gli esperti siamo noi)



→ seggiola



→ pedone

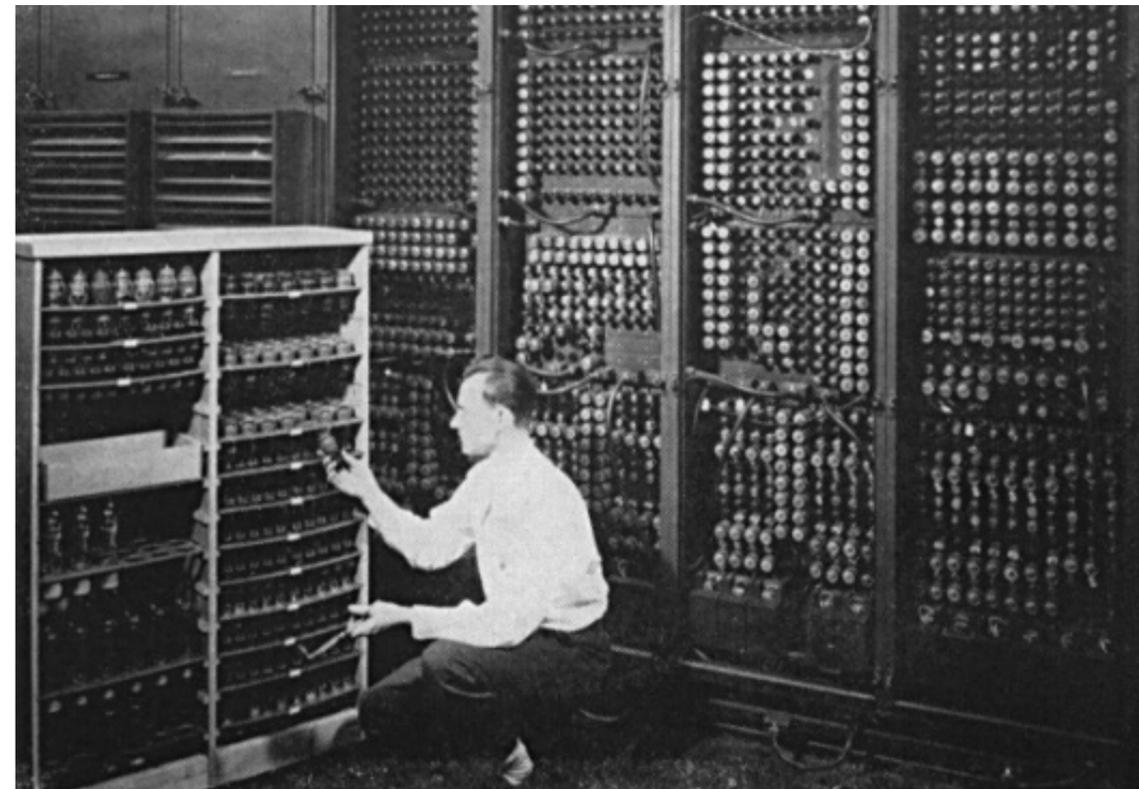


→ treno

Che cosa c'è dietro al recente successo?



I due veri protagonisti



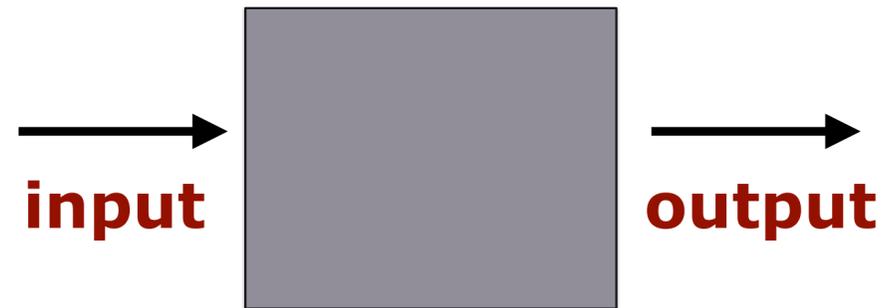
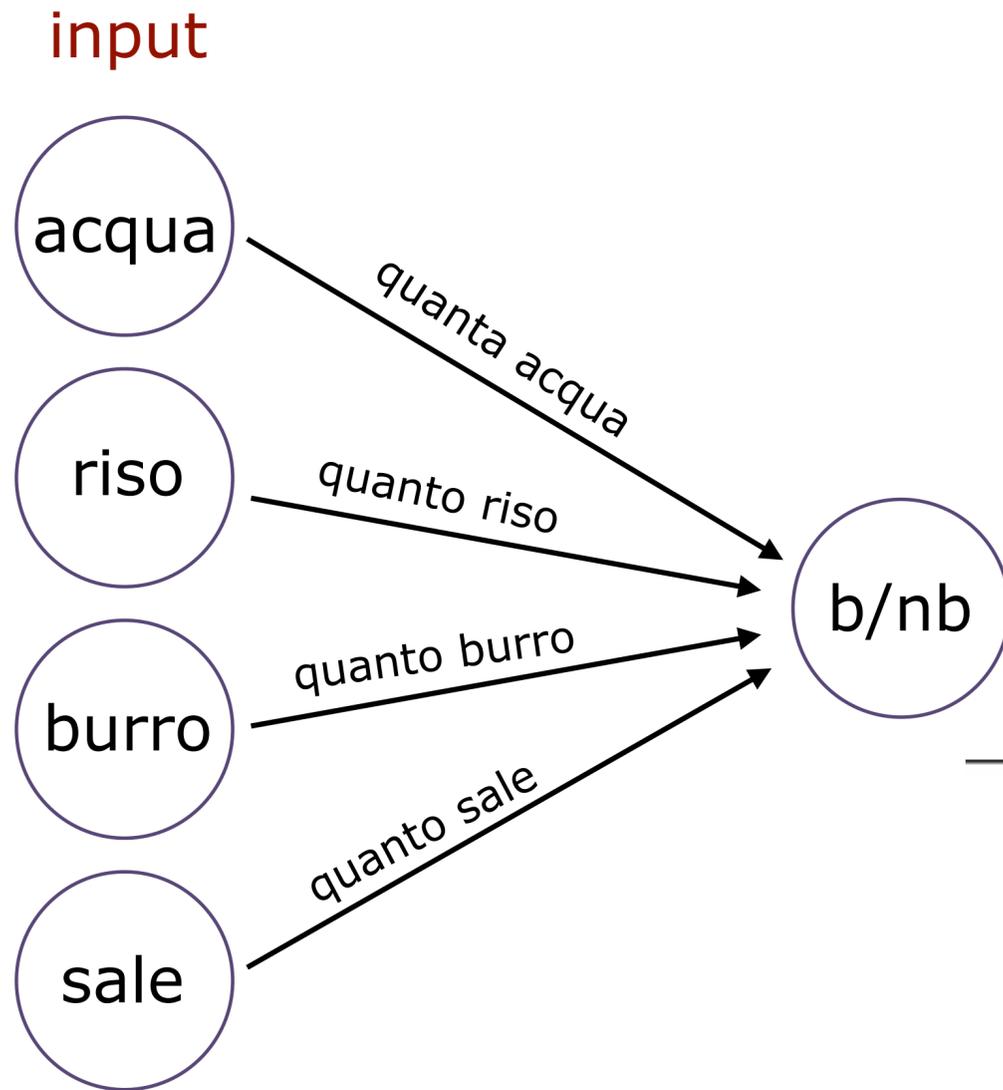
25 anni fa nel lab di Tommy Poggio



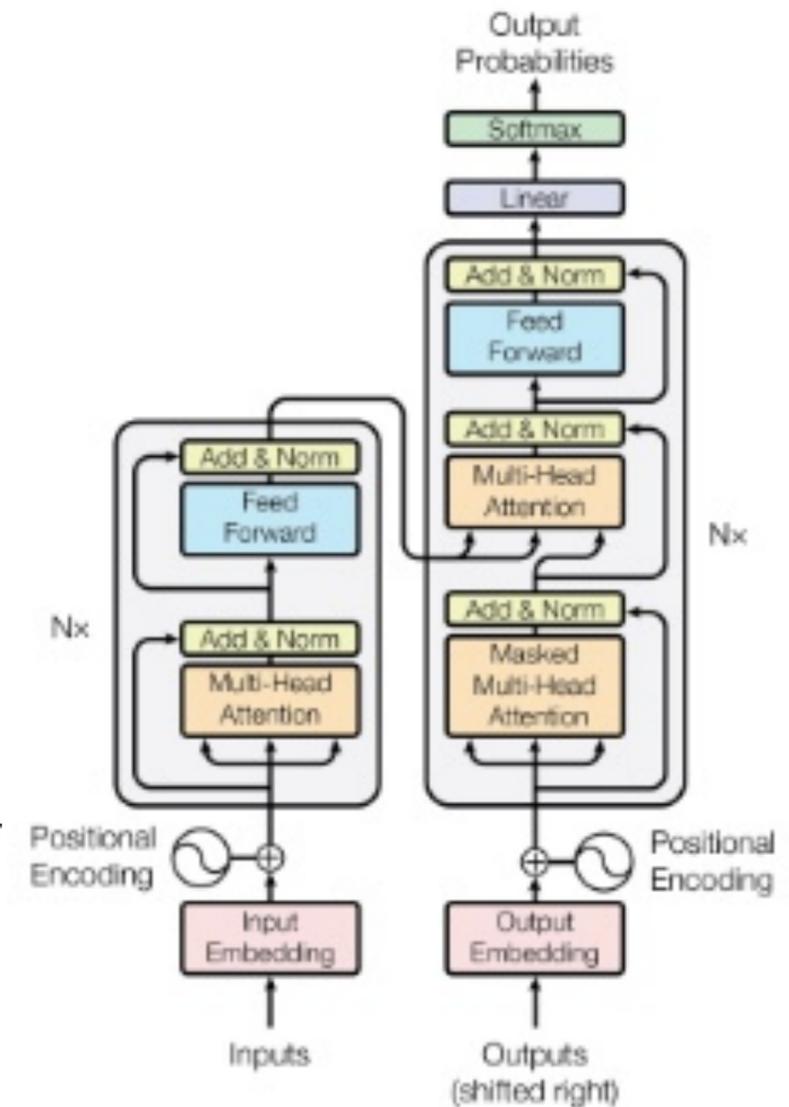
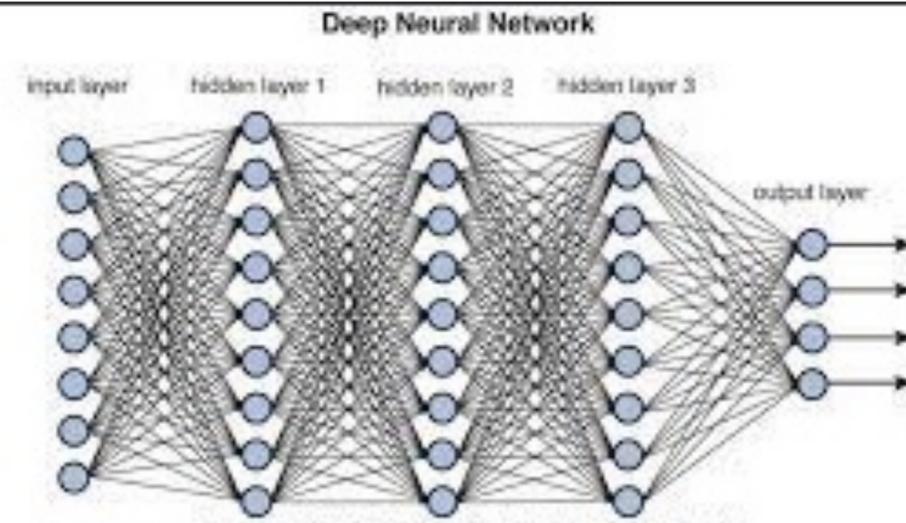
10 anni fa (Mobileye)



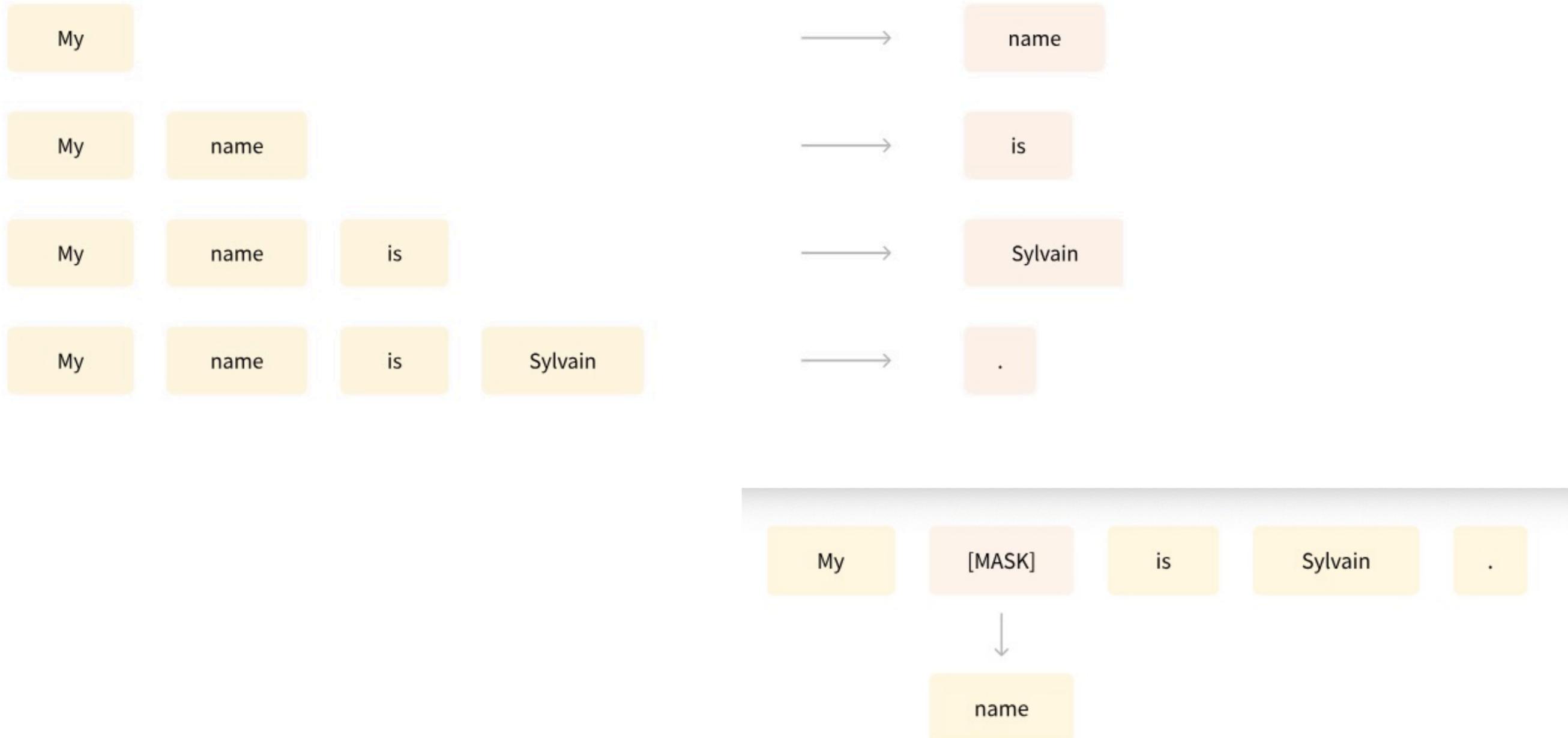
Neuroni dentro una scatola grigia



imparare = determinare i valori di circa 200,000,000,000 pesi



Esempi di esempi (senza esperti!)



L'AI è tra noi

ChatGPT 3.5 



You

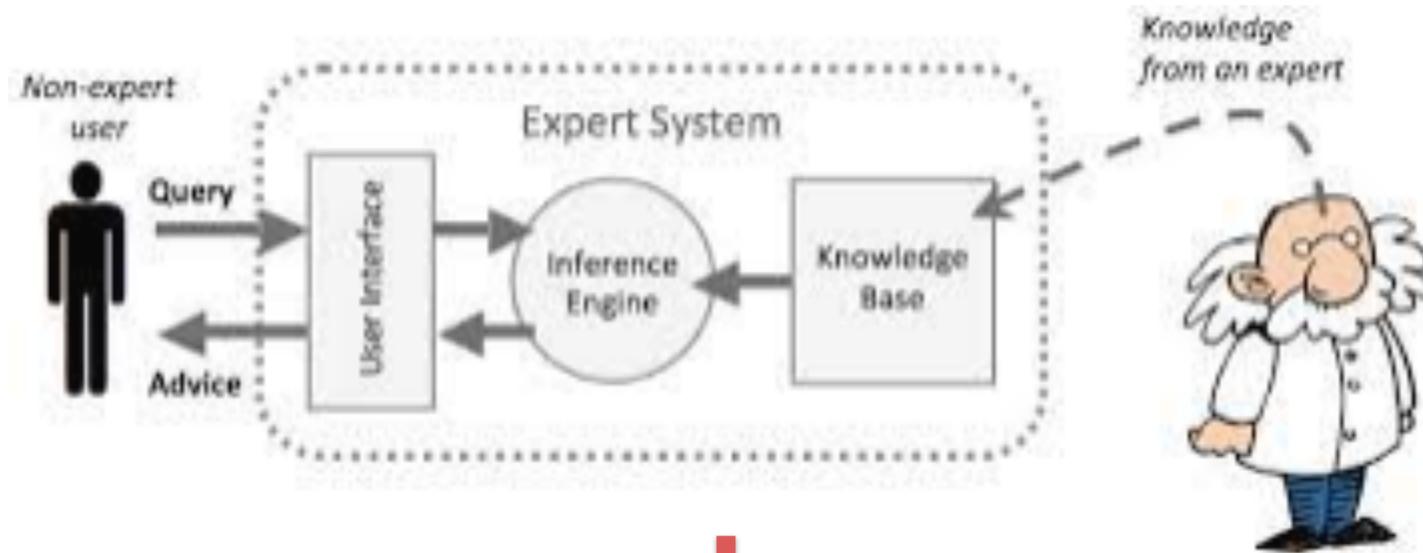
vorrei un breve riassunto di questa storia: un signore entra in un supermercato e compra una bottiglia di succo di frutta, una di vino, due pacchi di pasta, un etto di prosciutto, stracchino e una busta di spinaci. Si avvicina al banco surgelati e prende vongole e piselli. Si rende conto che deve interrompere brevemente per correre in bagno. Riesce ad arrivare in tempo e poi completa la spesa comprando detersivi per la casa, deodorante e schiuma di barba. Alla cassa paga in contanti e si allontana soddisfatto per lo sconto che ha ottenuto utilizzando la carta fedeltà.



ChatGPT

Certo! Un uomo va al supermercato e compra succo di frutta, vino, pasta, prosciutto, stracchino e spinaci. Mentre fa la spesa, si rende conto di dover andare in bagno e ci riesce in tempo. Poi completa gli acquisti con detersivi, deodorante e schiuma da barba. Alla cassa, paga in contanti e si sente contento per lo sconto ottenuto grazie alla carta fedeltà.

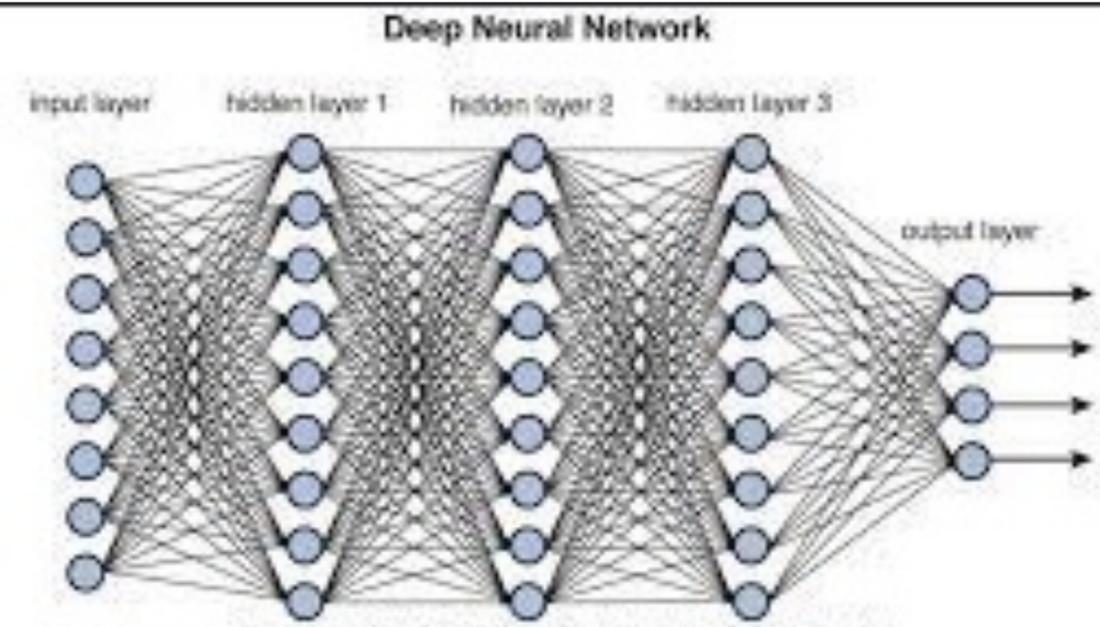
Due mondi a parte



Imparare dall'esperienza



Fare buon uso della conoscenza a priori per costruire rappresentazioni appropriate e modelli di ragionamento



Il diavolo è nei dettagli

- Come si giudica l'efficacia di una tecnologia di AI?
- E se i dati non sono tutto?
- È sufficiente cercare più dati o potrebbe mancare qualcosa?

Esempi di esempi (con esperti veri)

esami medici

→ diagnosi

andamento dei
mercati azionari

→ compra/vendi titolo

caratteristiche di un
appartamento

→ valutazione di mercato

acquisti effettuati

→ proposta di nuovo prodotto

Una tecnologia difficile da controllare

È possibile sviluppare un'intelligenza artificiale accessibile amichevole (friendly AI) per diminuire le diseguaglianze in modo che l'umanità possa trarne un effettivo beneficio?

Non è un ascensore

Se i tasti sono all'altezza giusta, se anche chi non vede o non sente bene è in grado di salire e scendere al piano desiderato, se l'accesso è ben studiato, se ...

Non è un sito internet

Se anche chi non vede o non sente bene è in grado di navigarlo, se la sua fruizione è semplice e intuitiva, se riporta informazioni accurate e controllabili, ...

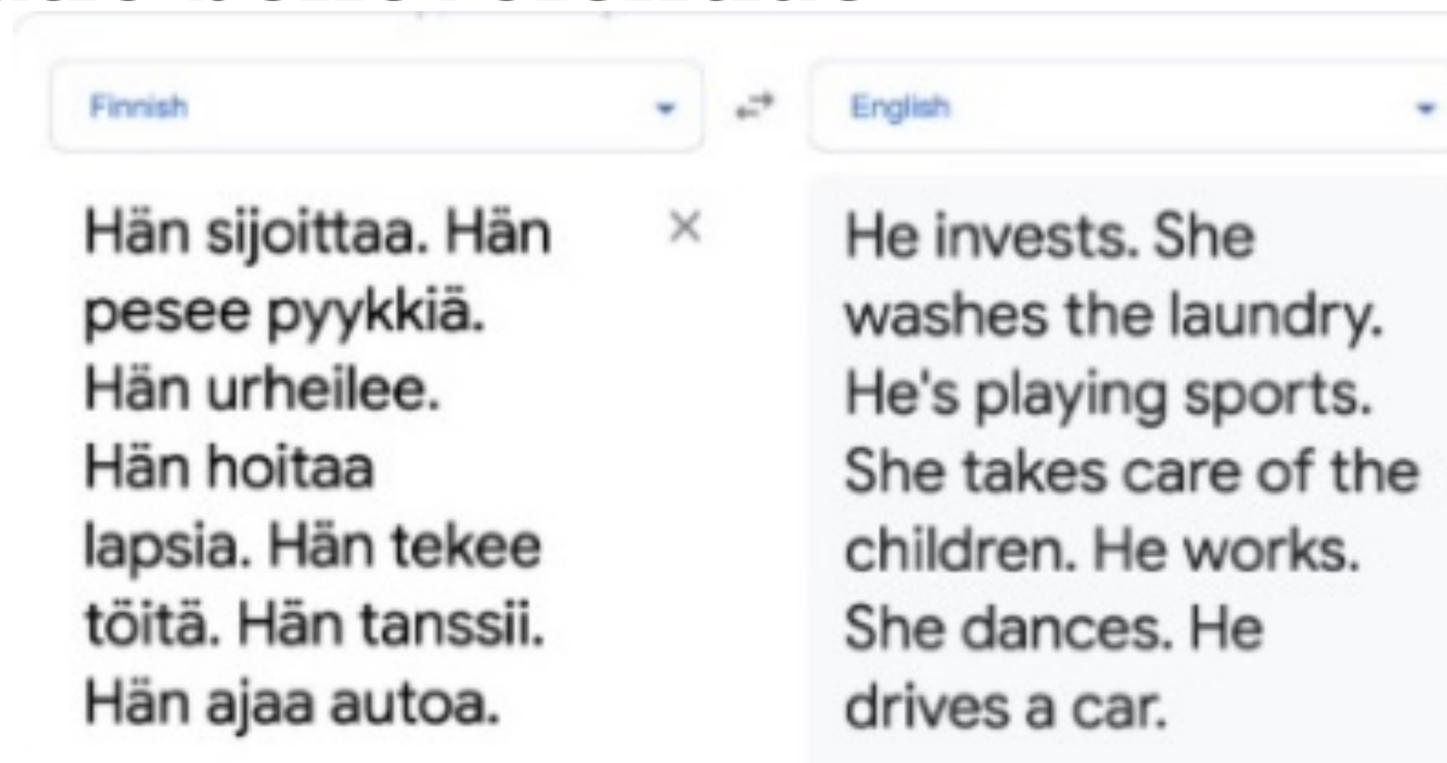
Categorie a rischio

Un sistema di AI (machine learning) è progettato per essere conformista

In presenza di altri dati e, soprattutto, di conoscenza specifica, è imperativo prendere in considerazione una decisione diversa!

Difficile da controllare

- La gentilezza di chatGPT è **captatio benevolentiae**
- Gli esempi sono (quasi sempre) intrisi dei nostri pregiudizi
- Non è necessariamente meglio aspettarsi che qualcuno decida per noi quali dati siano eticamente o, peggio ancora, moralmente accettabili



UniGe

MaLGA

The logo for MaLGA features a network graph with six white circular nodes and seven white edges. The nodes are arranged in a roughly circular pattern, with one node at the top, one at the bottom, and four nodes in between. The edges connect the nodes in a way that forms a central vertical line and several other connections, creating a complex network structure.