

UNA BREVE STORIA DELL'INTELLIGENZA ARTIFICIALE

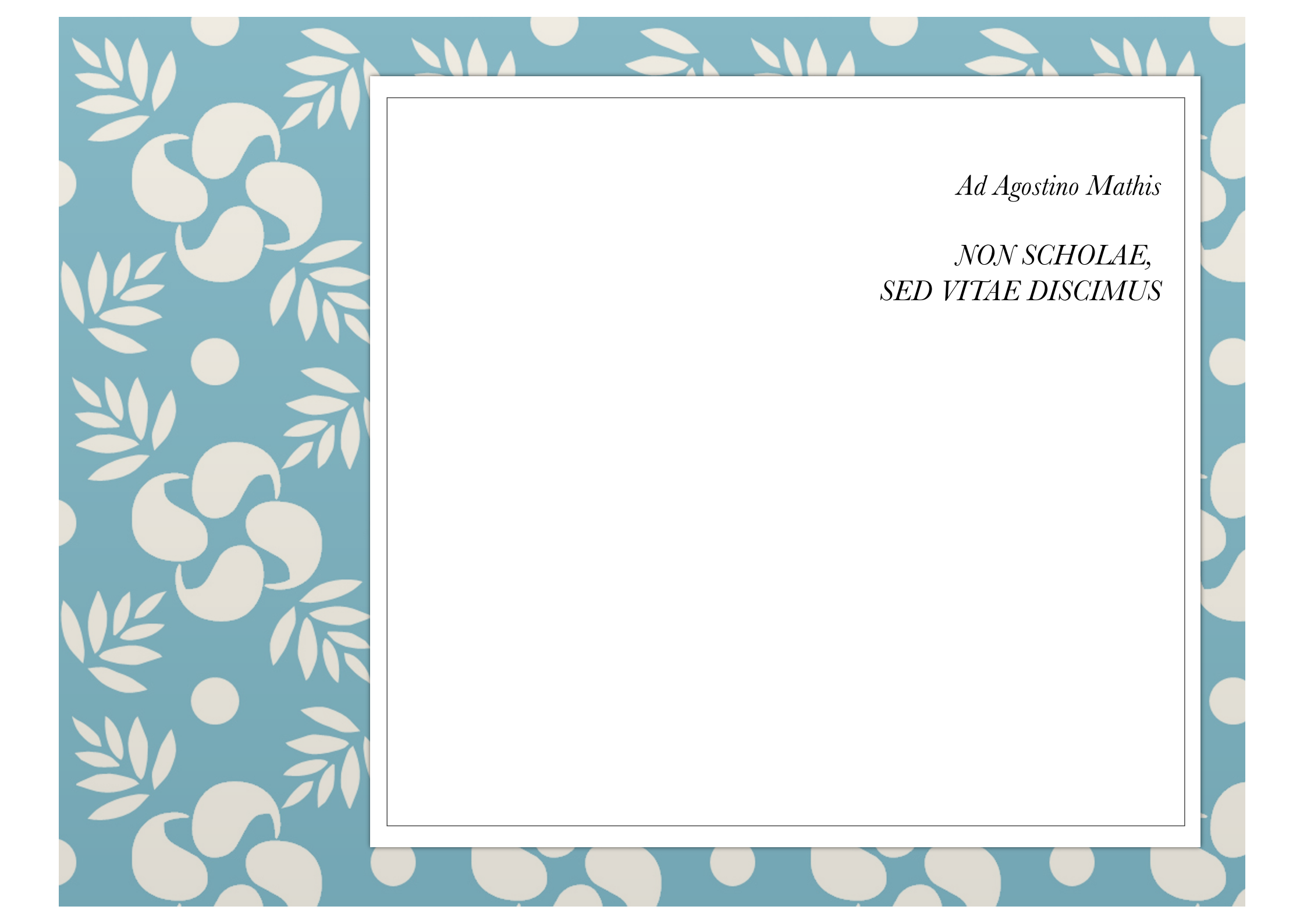
Ing. Mauro Cappelli
Commissione Ricerca e Reattori Innovativi



Ordine degli Ingegneri
della Provincia
di Roma

Dalle origini
filosofiche all'era
dei modelli
generativi con un
rapido sguardo a
possibili impieghi
nel nucleare

ROMA,
22 GENNAIO 2026



Ad Agostino Mathis

*NON SCHOLAE,
SED VITAE DISCIMUS*

Argomento

PARTE I: BREVE STORIA DELL'IA

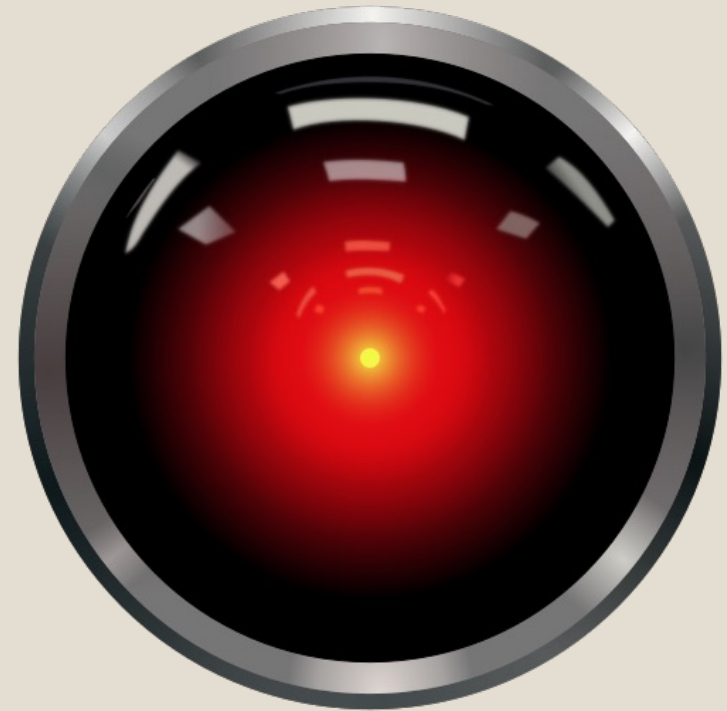
- Origini dell'IA
- Sviluppo storico
- Oggi e domani
- Impatto culturale e questioni etiche

PARTE II: POSSIBILI IMPIEGHI IN AMBITO NUCLEARE

- Uno sguardo internazionale
- Cosa aspettarci (cosa no)

Prolegomenon

- Di cosa si parla
 - «essere stato» è preconditione dell'«essere»
- Di cosa NON si parla
- Limiti di tale presentazione
 - Spazio negli altri interventi
 - Distacco critico



L'occhio rosso di HAL 9000
(2001: Odissea nello spazio, Stanley Kubrick, 1968)

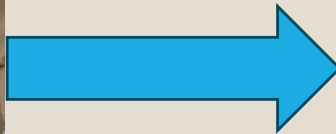
Principi guida

«L'incomprensione del presente nasce fatalmente dall'ignoranza del passato.» **Marc Bloch**

«L'evento è polvere: solo la lunga durata fa la storia.» **Fernand Braudel**

«La storia nasce dallo scarto tra spazio di esperienza e orizzonte di attesa.» **Reinhart Koselleck**

Obiettivo del seminario



Cos'è l'IA



Che cos'è l'Intelligenza?

«Qualità mentale che consiste nella abilità di apprendere dall'esperienza, di adattarsi a nuove situazioni, comprendere e gestire concetti astratti, e utilizzare conoscenza per agire sul proprio ambiente». (Enciclopedia Britannica)



IA: Simulazione dell'Intelligenza

- Comprensione
- Memorizzazione
- Riproduzione

Lo studio di come rendere i calcolatori capaci di pensare nel vero senso letterale del termine (Haugeland, 1985).

→ realizzare sistemi che pensano come gli uomini.

Lo studio di come far fare ai calcolatori cose che, ora come ora, gli uomini fanno meglio (Rich and Knight, 1991).

→ realizzare sistemi che agiscono come gli uomini.

Lo studio di come spiegare ed emulare il comportamento intelligente attraverso processi computazionali (Schalkoff, 1990).

→ realizzare sistemi che agiscono razionalmente.

È lo studio delle facoltà mentali attraverso l'uso di modelli computazionali (Charniak and McDermott, 1985).

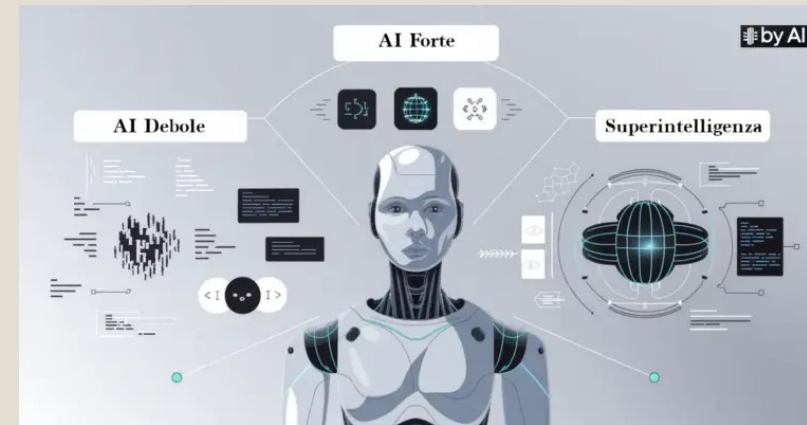
Cioè realizzare sistemi che pensano razionalmente.

Definizione **condivisa** di Intelligenza Artificiale

«Branca dell'informatica che studia le proprietà dell'intelligenza **sintetizzando** intelligenza»

Circolare, ma anche aperta a qualsiasi sviluppo. Cosa che può vedersi come negativa ma che d'altro canto ha portato a una miriade di sviluppi impensabili.

IA debole e forte



- Compiti **specifici** vs Intelligenza **generale** → lungo dibattito filosofico

IA forte o di livello generale:

obiettivo: comprensione e simulazione dell'intelligenza umana

metodo: costruzione di modelli computazionali plausibili della mente umana e dei suoi processi

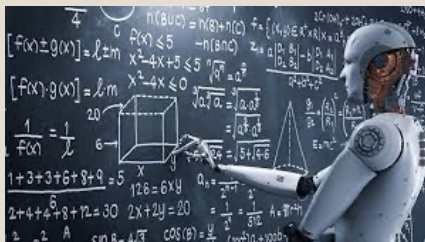
validazione: risolvere i problemi con gli stessi processi usati dall'uomo

IA debole o di livello specifico:

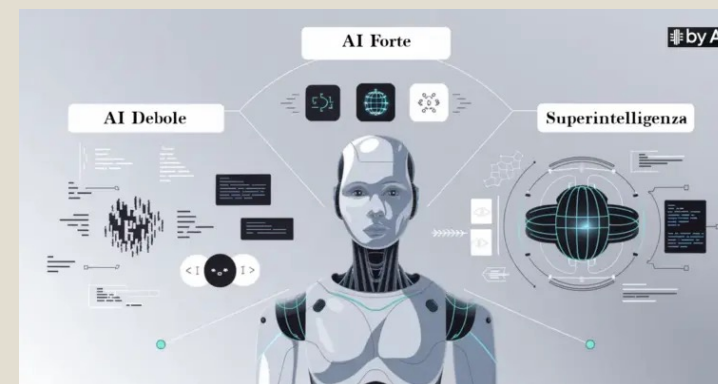
obiettivo: costruzione di entità dotate di intelligenza razionalizzata o razionalità

metodo: logico classico, i.e. codifica del pensiero razionale

validazione: l'importante è risolvere i problemi che richiedono intelligenza (più pratico e utilitaristico)

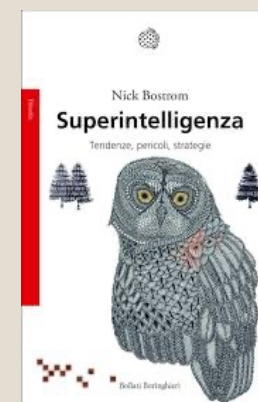


Superintelligenza



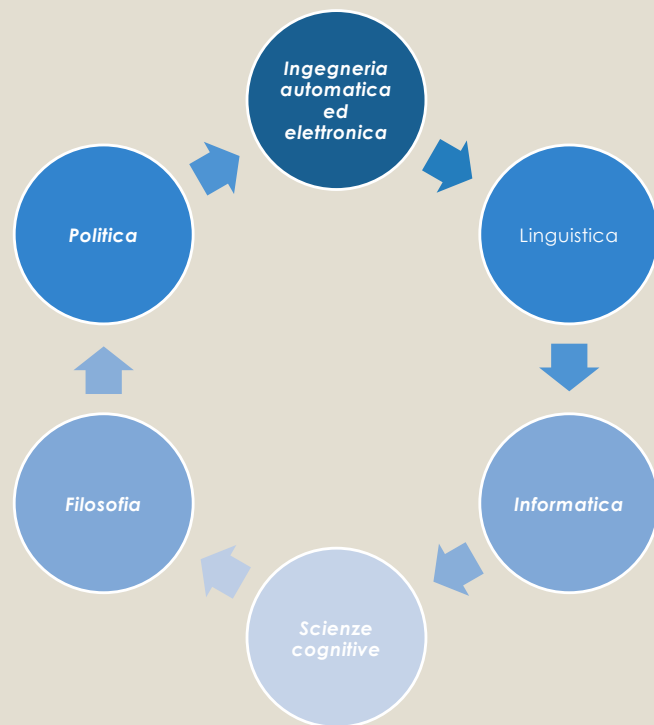
Superintelligenza

Sistema ipotetico di AI basato su software con una portata intellettuale che va *oltre l'intelligenza umana*. Al livello più fondamentale, questa AI superintelligente possiede funzioni cognitive all'avanguardia e skill di pensiero più avanzate di quelle di qualsiasi essere umano.



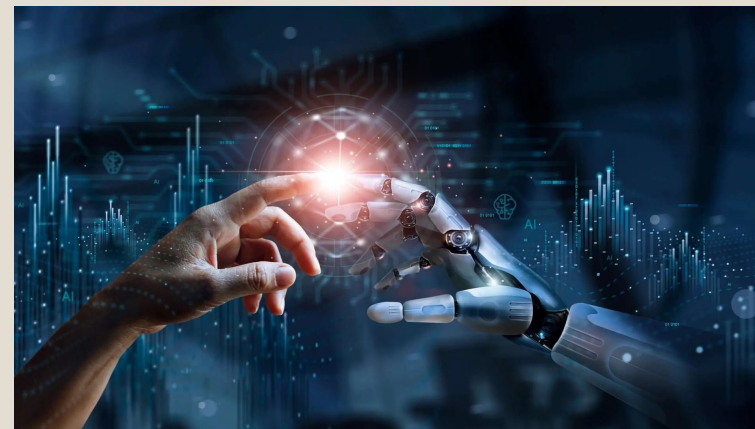
Lo studio dell'IA

DISCIPLINE

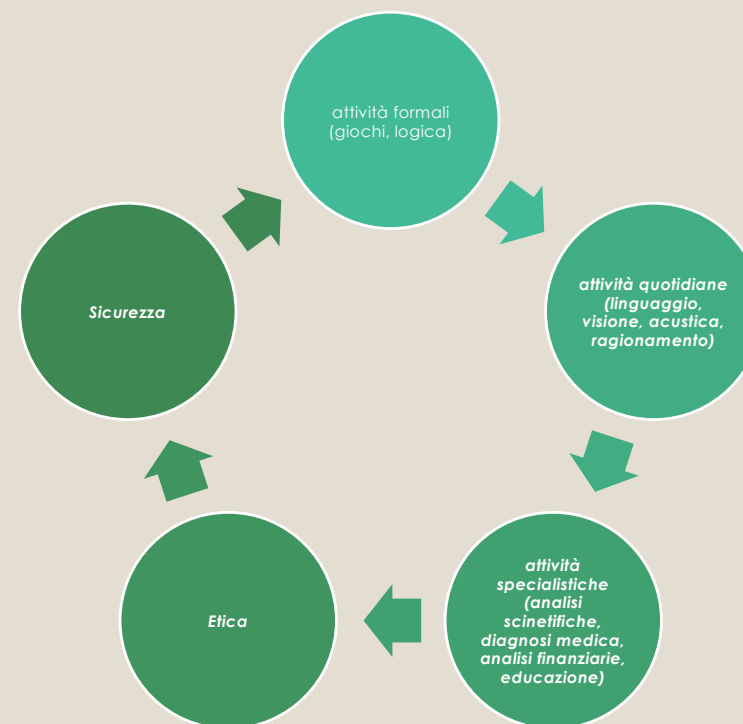


È dunque un problema più *scientifico* (replicare il cervello umano) o *tecnologico* realizzare machine automatiche)?

"Dipende da chi mette i soldi..."



PROBLEMI



Radici filosofiche

L'IA è un campo di ricerca relativamente giovane (possiamo fissare la sua nascita negli anni Cinquanta), ma affonda le sue radici nella filosofia antica e moderna fino alla filosofia contemporanea che ha approfondito i seguenti temi:

- Logica
- Razionalismo
- Formalismo

Accanto ai filosofi (i primi a interrogarsi su cosa sia la *comprensione* e la *conoscenza*) sono via via emersi altri studiosi dalle nuove scienze.



Platone illustra l'idea dei solidi platonici
(immagine elaborata con Midjourney)

Platone e Aristotele: ridurre il ragionamento a calcolo meccanico

Platone descrive in un suo dialogo Socrate che cerca un modo per distinguere ciò che è «pietà» da ciò che è «non pietà» (concetto di **algoritmo** *in nuce*)

Aristotele cerca di formulare le leggi che governano la parte razionale della mente e getta le basi di:

- Logica sillogistica
- Inferenza
- Ragionamento formale

Aristotele si può considerare il padre
della logica e del ragionamento
algoritmico



Raffaello, Scuola di Atene, Musei Vaticani

Il Medioevo: luci nell'oscurità

Nel **Medioevo** (epoca da rivalutare) si introducono:

- Automi
- Sapere tecnico
- Meccanizzazione

Al-Jazari nel XIII secolo propose un **automa programmabile**, aprendo così la strada per lo sviluppo di una seminale **robotica**. Scrisse un intero libro con descrisse cinquanta dispositivi meccanici (**automi**) con le istruzioni per costruirli.



L'automa per l'abluzione delle mani

Età moderna



Giordano Bruno e **Raimondo Lullo** introducono una forma razionale allo studio dei fenomeni della natura:

- Ars combinatoria e mnemotecnica
- Calcolo della verità
- Pre-IA simbolica

Francis Bacon con il “*Novum Organum*” dà vita al movimento dell’*empirismo*.

In generale, l’Era Moderna approfondisce le tematiche:

- Meccanicismo
- Natura-macchina
- Razionalismo

Cartesio, Leibnitz, Hume

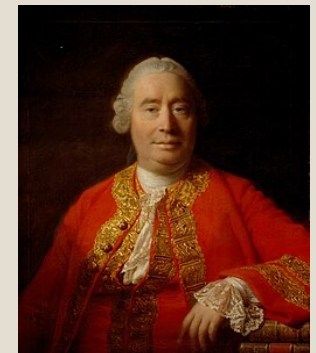
Con **Cartesio** fa la comparsa un concetto che segnerà tutto il pensiero successivo, il **dualismo** (anima-corpo):

- Animali-macchina
- Linguaggio
- Limiti delle macchine

Leibnitz formula esplicitamente il sogno dell'automazione del ragionamento:

- *Calculus ratiocinator*
- Lingua universale
- Calcolo del pensiero

David Hume nel suo capolavoro “*A Treatise of Human Nature*” propone formalmente il principio dell'**induzione**



XIX-XX secolo

Nell'**Ottocento**, la Rivoluzione industriale aumenta i sogni su nuove *macchine* dotate di automazione.

Charles Babbage propone il primo prototipo di calcolatore (*macchina differenziale*):

- Macchina analitica
- Programma
- Calcolo



Ada Lovelace (figlia legittima di Lord Byron) contribuisce alla macchina di Babbage con un algoritmo per generare i numeri di Bernoulli, considerato come **il primo algoritmo** espressamente inteso per essere elaborato da una macchina (Ada Lovelace è spesso ricordata come la prima programmatrice di computer al mondo).



Logica matematica



- **George Boole** (1815-1864) introduce un *linguaggio formale* per formalizzare le leggi dell'inferenza logica
- **Gottlob Frege** (1848-1925) definisce una *logica* che, eccetto alcune modifiche di notazione, è la logica del prim'ordine usata in gran parte dei sistemi odierni di rappresentazione della conoscenza.
- **Bertrand Russell** (1872-1970) è un esponente del *positivismo logico* e descrive i *fondamenti della matematica*.

Cominciano a definirsi concetti chiave per la rappresentazione della conoscenza (*epistemologia*): Linguaggio formale, Logica del predicato, Significato.

- **Alfred Tarski** (1902-1983) introduce una *teoria* che permette di mettere in relazione gli oggetti del mondo reale con gli oggetti di una rappresentazione logica
- **Kurt Gödel** (1906-1978) prova il cosiddetto *problema dell'indecidibilità* e formula i due *teoremi di incompletezza* che affossano l'idea di Hilbert di voler costruire un *formalismo matematico universale* (Gödel mostra che in qualsiasi sistema formale non è mai possibile giungere a definire la lista completa degli assiomi che permetta di dimostrare tutte le verità)

Alan Turing (1912-1954)



Il problema della Logica si sposta sul terreno dell'**Informatica**:

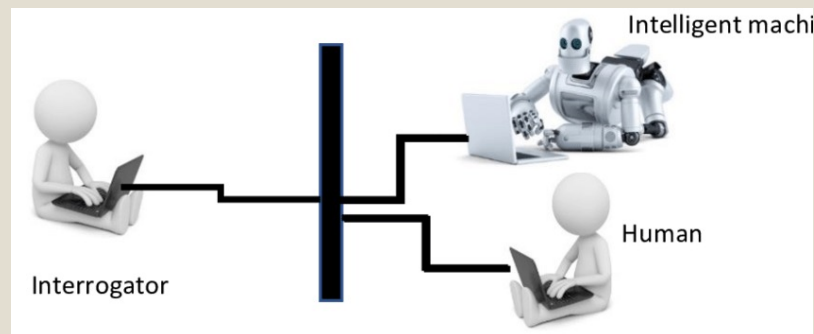
- Macchina di Turing
- Calcolabilità
- Algoritmi

Nasce l'**Ingegneria dei calcolatori**:

- Nel 1940 in Inghilterra venne costruito dal gruppo di Alan Turing il primo computer moderno, **Heath Robinson**
- Nel 1941 in Germania venne costruito da Konrad Zuse il primo computer programmabile, **Z-3**
- Tra il 1940 e il 1942 negli Stati Uniti venne realizzato il primo calcolatore elettronico, **ABC**
- Nel 1945 viene realizzato per Z-3, il **primo linguaggio di programmazione ad alto livello**

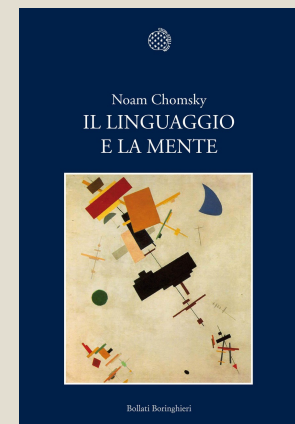
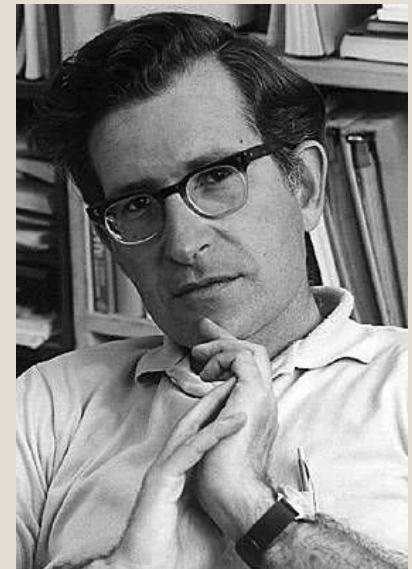
Test di Turing

- Il Test di Turing, proposto da Alan Turing nel 1950, è un esperimento per valutare se una macchina possa esibire un comportamento intelligente *indistinguibile da quello umano*
- È basato su una conversazione testuale tra un valutatore umano, un umano e un computer, con l'obiettivo di capire se l'interrogatore non riesca a distinguere il computer dall'essere umano.
- In sostanza, se una macchina riesce a ingannare un umano facendogli credere di essere un altro essere umano, supera il test, dimostrando una capacità di imitazione del comportamento umano sufficiente per essere considerata "intelligente" in quel contesto.



Contributi da psicologia, biologia e linguistica

- **Herman von Helmholtz** (1821-1894) e **Wilhelm Wundt** (1832-1920) applicano il metodo scientifico allo studio della visione umana
- **John Watson** (1878-1958) e **Edward Lee Thorndike** (1874-1949) con il loro movimento del **comportamentismo** basano le loro teorie sui concetti di stimolo e risposta e rifiutano ogni teoria basata su processi mentali
- **Warren McCulloch** e **Walter Pitts** nel 1943 propongono un modello di *neurone artificiale*
- Nel 1957 **Noam Chomsky** descrive la teoria delle strutture sintattiche
- I modelli sintattici di Chomsky sono abbastanza formali da poter essere programmati.



Dartmouth 1956: nasce l'IA

- Nel 1956 durante il **workshop di Dartmouth** John McCarthy coniò per questi tipi di ricerca il nome di **Intelligenza Artificiale**

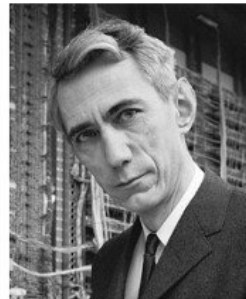
1956 Dartmouth Conference: The Founding Fathers of AI



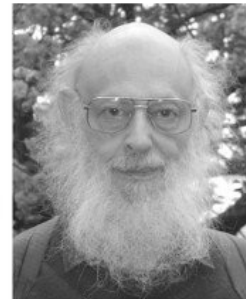
John MacCarthy



Marvin Minsky



Claude Shannon



Ray Solomonoff



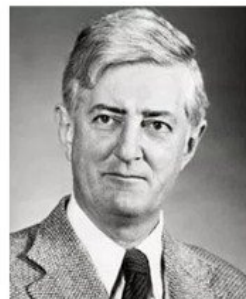
Allen Newell



Herbert Simon



Arthur Samuel



Oliver Selfridge



Nathaniel Rochester



Trenchard More

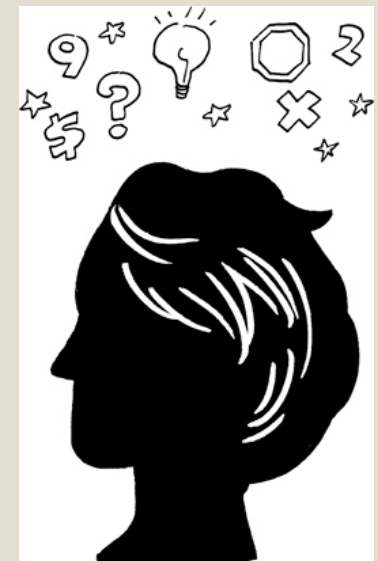
Dartmouth 1956: nasce l'IA

Risultati del meeting

- Furono sanciti i quattro pionieri: John McCarthy, Marvin Minsky, Allen Newell e Herbert Simon. Tutti si dissero sicuri che in capo a massimo due decenni tutti i problemi della realizzazione dell'IA si sarebbero risolti!
- Fu scelto il nome definitivo di **Artificial Intelligence**, senza troppa convinzione da **McCarthy** (l'obiezione era che l'intelligenza dovesse essere “autentica”!).
- Nella proposta i quattro scrivevano che l'ipotesi da cui partivano era che ogni caratteristica dell'intelligenza fosse “descrivibile con precisione tale da poter costruire una macchina capace di simularla”. La cosa sorprendente è che i punti chiave descritti nella proposta sono gli stessi su cui ci si arrovela tuttora:
 - elaborazione del linguaggio naturale,
 - reti neurali,
 - apprendimento automatico (machine learning),
 - astrazione e ragionamento,
 - creatività.

I primi anni

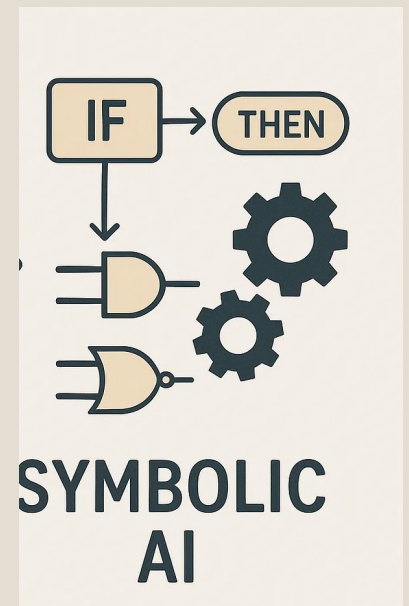
- John McCarthy nel 1958 definisce il **LISP** (solo negli Ottanta saranno sviluppati dei calcolatori, LISP Machine, con processori orientati all'esecuzione di programmi LISP)
- **Allen Newell** and **Herbert Simon** nel 1956 realizzano un programma di ragionamento, *Logic Theorist* e un programma per la risoluzione di problemi, *General Problem Solver* (GPS).
- **GPS** è il primo *computer program* concepito come **macchina universale risoltrice di problemi**: lo stesso programma ha l'ambizione di risolvere con lo stesso algoritmo ogni tipo di problema.
- Newell e Simon sostenevano che creare un numero sufficiente di queste connessioni avrebbe reso le macchine più simili a noi e che una parte fondamentale del ragionamento umano consistesse semplicemente nel collegare **simboli**: il nostro linguaggio, le nostre idee e i nostri concetti sono semplicemente ampi raggruppamenti di **simboli interconnessi**.
- Ha una struttura interna di tipo *search* che impone alti costi computazionali e di tempo per risolvere problemi di tipo realistico.



IA simbolica

- L'IA simbolica genera conoscenza tramite simboli (parole o frasi) in un linguaggio comprensibile all'uomo e delle regole per la loro combinazione ed elaborazione.
- Cosa significa per un computer un simbolo? Naturalmente il computer non lo interpreta come un oggetto dotato di senso ma attribuisce un significato ai simboli in base a come essi sono combinati.
- Chi sostiene l'IA simbolica ritiene che non sia necessario costruire programmi che simulano il cervello ma l'intelligenza emergerebbe dalle relazioni tra simboli. L'evoluzione di tali idee fu chiamato **sistema esperto**.
- Gli umani progettavano le regole che poi i calcolatori applicavano nelle decisioni (diagnosi clinica, decisioni giuridiche). Un esempio storico di IA simbolica fu il General Problem Solver (GPS) di Simon e Newell.

Tale approccio **simbolico** è stato messo in discussione tra gli altri da **John Searle**



L'argomento della Stanza Cinese

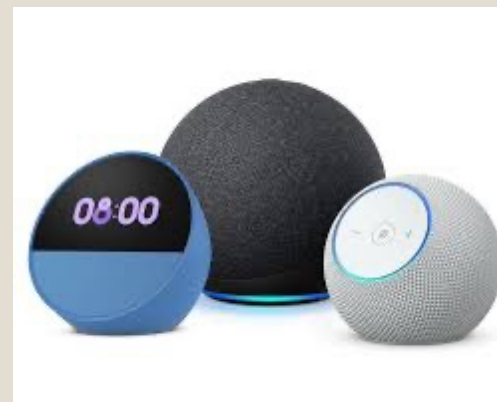
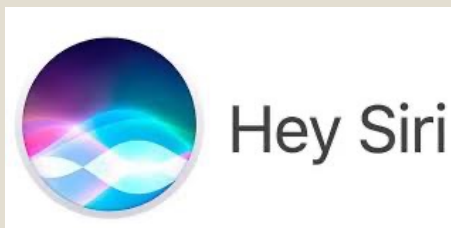
- Nel 1980, il filosofo **John Searle** sostenne che la semplice connessione di simboli non potesse essere considerata intelligenza. Per sostenere la sua argomentazione contro l'affermazione che i computer pensino o almeno abbiano il potenziale per farlo un giorno, creò un esperimento chiamato "**argomento della stanza cinese**".

L'esperimento della *stanza cinese* di John Searle immagina una persona che parla solo inglese chiusa in una stanza, capace di rispondere a messaggi in cinese seguendo istruzioni meccaniche, senza capire nulla di ciò che legge o scrive. Per chi sta fuori, le risposte sembrano intelligenti, ma in realtà non c'è alcuna comprensione. Searle usa questo esempio per sostenere che la semplice manipolazione di simboli non equivale a intelligenza.



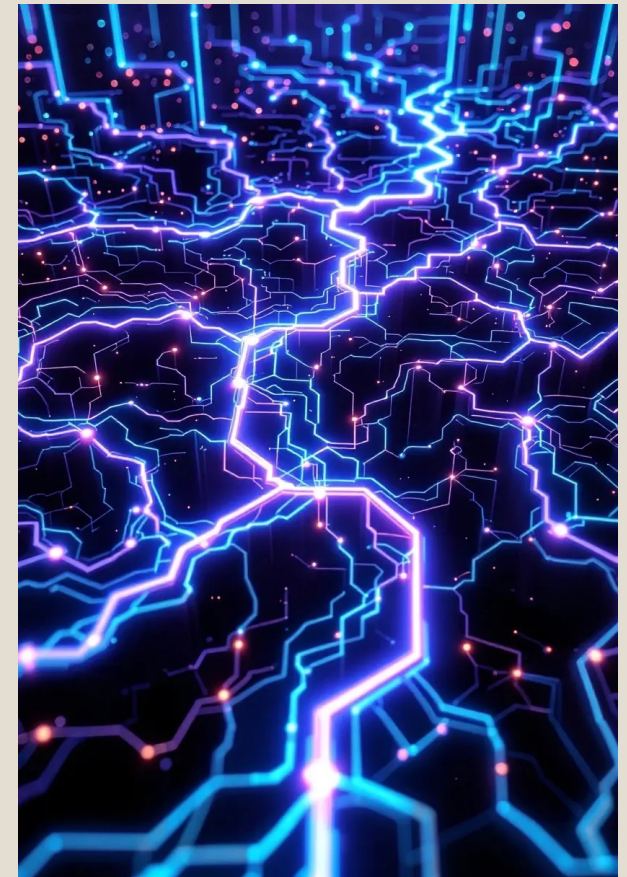
L'argomento della Stanza Cinese

- Lo stesso vale per assistenti vocali come **Siri** o Alexa, che rispondono a domande senza capirne il significato.
- Un limite fondamentale di questo approccio è la “esplosione combinatoria”: il numero di possibili domande e risposte cresce rapidamente, rendendo il sistema inefficiente.
- Per circa 25 anni l'IA si è basata sul confronto di simboli, ma questo metodo non è riuscito a gestire la complessità crescente, risultando lento e inadeguato.



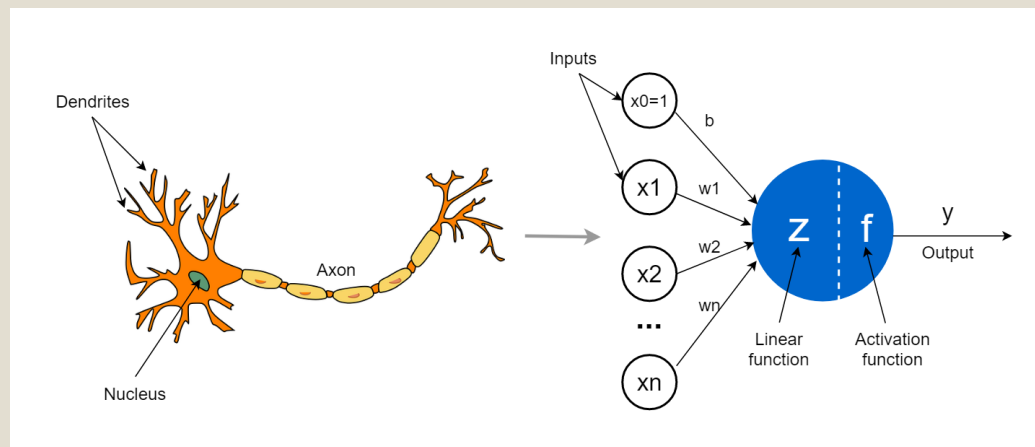
IA subsimbolica

- L'IA subsimbolica genera invece conoscenza non tramite un linguaggio simbolico cosciente o consapevole ma tramite processi mentali inconsapevoli.
- Se l'IA simbolica fa riferimento a come le persone riferiscono di pensare usando un linguaggio logico-matematico, l'IA subsimbolica fa riferimento alle **neuroscienze** e allo studio dei processi inconsci (per es. come riconoscere volti e identificare parole).
- Il linguaggio non è quello umano ma quello matematico delle equazioni che trasformano dati in regole (cfr. equazione dei pesi di un percettrone).



Percettroni

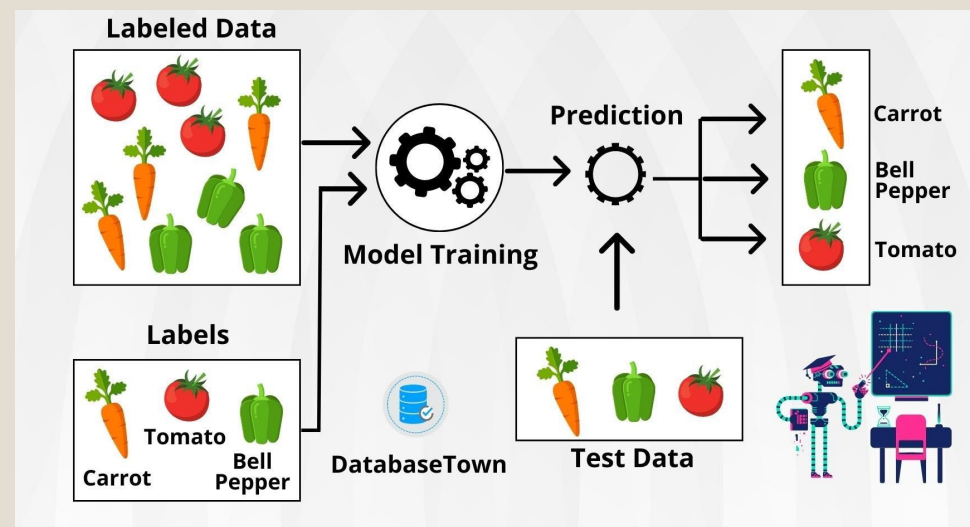
- Il primo esempio di IA subsimbolica è il **perceptrone** proposto dallo psicologo Frank Rosenblatt, antesignano delle attuali deep neural networks.
- Il perceptrone si ispira al neurone di cui ricalca il funzionamento.
- Analogamente a un neurone, un perceptrone scarica (output=1) se la somma dei suoi input ponderati supera la soglia, altrimenti no (output=0): tutta la conoscenza è nella codifica di pesi e soglia.
- Es.: noi decidiamo così sull'andare a vedere un film, se alcuni amici di cui ci fidiamo ce lo raccomandano il loro giudizio vale più di quello di altri.



Percettroni: apprendimento

- Come fa il percettrone a imparare come deve comportarsi?
- Secondo Rosenblatt, il percettrone deve imparare tali valori da solo, come un normale cervello umano, sulla base del **condizionamento** (all'epoca era in volga la psicologia comportamentista di Burrhus Frederic Skinner): ricevere un premio se svolge bene il compito, una punizione in caso contrario (*piccioni di Skinner*).

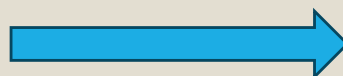
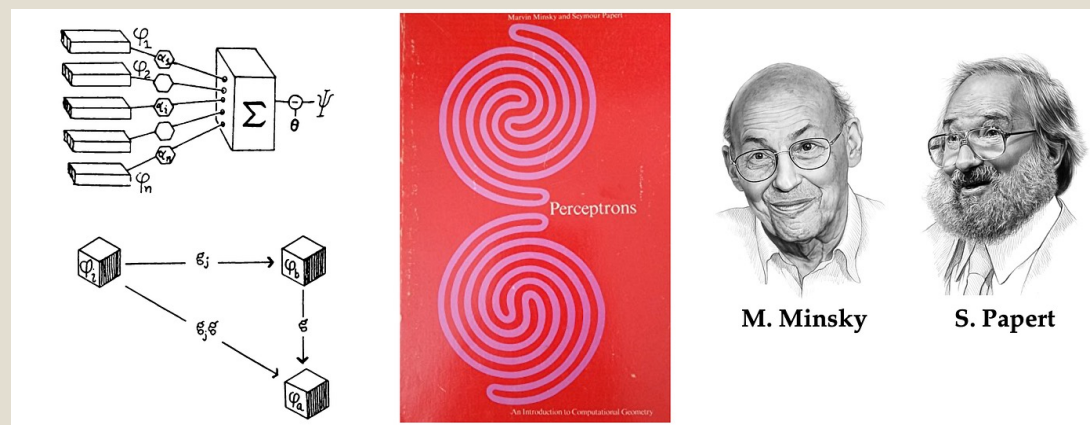
Tale IA è nota oggi come **supervised learning** (apprendimento supervisionato), in quanto riceve un voto sulla differenza tra output prodotto e output desiderato, che poi usa per ridefinire pesi e soglia.



Gli anni Sessanta: primo AI winter

- Rosenblatt mostrò che i percettroni erano in grado di risolvere problemi di riconoscimento semplici.
- Ma nel 1969 **Marvin Minsky e Seymour Papert** nel loro libro *Perceptrons* mettono in luce i limiti delle reti neurali basati su percettroni! Il percettrone era perfetto con problemi semplici ma andava in crisi in presenza di **molte pesi e soglie**.
- Per aumentarne le potenzialità occorreva aggiungere uno strato di neuroni simulati a formare una **rete neurale multistrato**, oggi la base della moderna IA, ma all'epoca poco studiate e difficili da addestrare.

Pertanto, le linee di ricerca sui percettroni e l'AI subsimbolica non furono più finanziate: l'IA entrò nel suo **primo «inverno»**. I finanziamenti andavano e venivano così come le mirabolanti promesse....



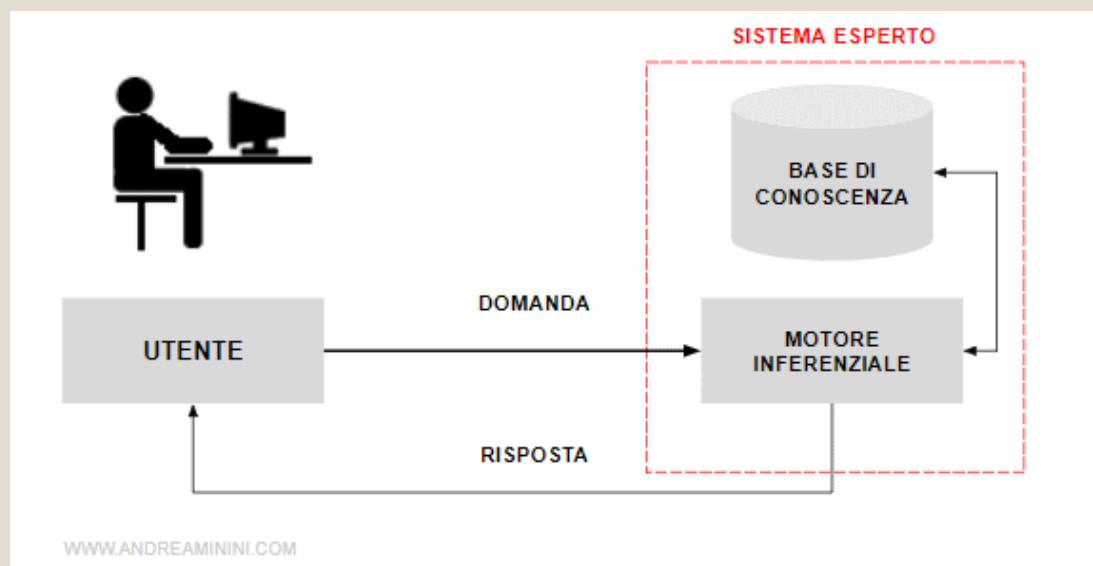
L'IA si divide in due strade!

Torna l'IA simbolica: Sistemi esperti

Un **sistema esperto** è un programma che cerca di riprodurre le prestazioni di una o più persone esperte in un determinato campo di attività

I programmi utilizzati dai sistemi esperti sono in grado di porre in atto procedure di **inferenza** adeguate alla risoluzione di problemi particolarmente complessi

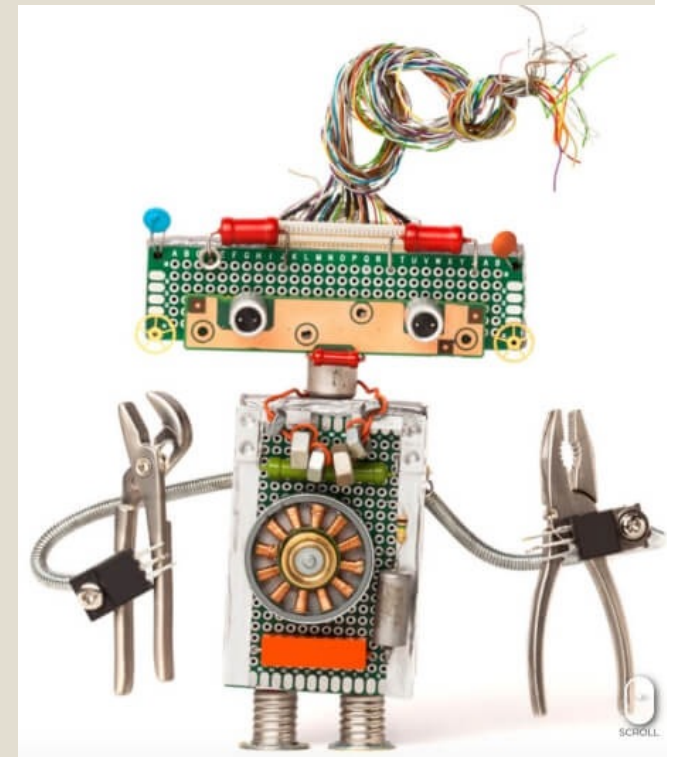
- Nel 1969 il gruppo di Bruce Buchanan e Edward Feigenbaum realizza il primo **sistema esperto**, **DENDRAL** (analisi chimica). Nel 1976 lo stesso gruppo realizza **MYCIN** un sistema esperto per la diagnosi di infezioni del sangue



Sistemi esperti: crisi e declino

Negli anni '70 e '80, i sistemi esperti furono salutati come la rivoluzione dell'intelligenza artificiale, promettendo di automatizzare decisioni complesse in settori come la medicina, la finanza e l'ingegneria.

- Tuttavia, nonostante gli investimenti massicci e l'entusiasmo iniziale, questi sistemi non riuscirono a mantenere le aspettative per i seguenti motivi:
 - Fragilità della Conoscenza Codificata: Difficoltà di Aggiornamento, Ambiguità e Incertezza, Scalabilità Limitata
 - Problema del “Bottleneck della Conoscenza”: Lentezza e Costo, Soggetto a Errori, Non Scalabilità
 - Mancanza di Flessibilità e Adattabilità: rigidità, difficoltà di integrazione, costi di manutenzione
 - Aspettative irrealistiche: Sostituire gli Esperti Umani, Risolvere Problemi Complessi, Essere Universali
 - Ascesa di Alternative Più Efficaci: Machine Learning, Big Data, GPU



incapacità di gestire l'incertezza e rigidità delle regole

Secondo AI Winter

Negli anni Ottanta l'IA è di nuovo in crisi.

- Gli stessi pionieri se ne accorsero:
 - “l'IA era un obiettivo più difficile di quanto pensavamo” (John McCarthy);
 - “Le cose facili sono difficili (**Easy things are hard**)” (Marvin Minsky).
- In effetti l'IA vive un paradosso: sa fare cose molto complesse come vincere a scacchi ma non sa fare bene le cose semplici come le farebbe un bambino (p.e. parlare o scrivere frasi poetiche).



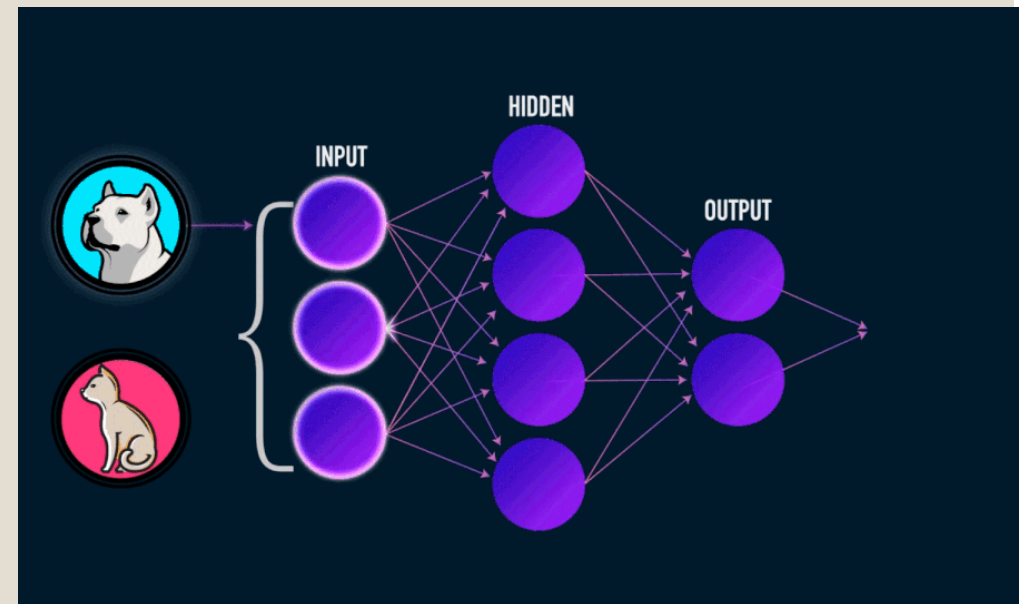
Machine Learning: un cambio di paradigma

Lentamente però il crescere di:

- Disponibilità di molti *dati* (avvento del web)
- Miglioramento delle tecniche di *apprendimento*
- *Piattaforme tecnologiche* sempre più potenti

hanno rilanciato l'IA subsimbolica centrandola sul *Machine Learning* tanto che oggi è diventato una metonimia per IA!

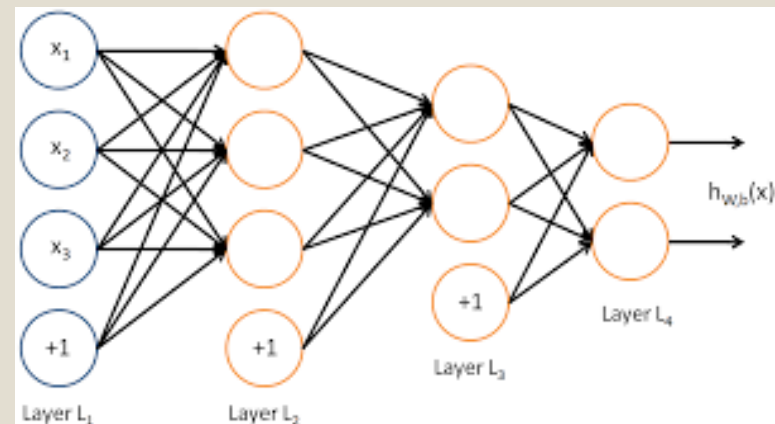
Ciò è dovuto in breve al fatto che il risultato più di successo degli ultimi anni è quello delle **reti neurali multistrato** (**ML-NN**, **Multi-Layer Neural Networks**), un'estensione dei percettroni, sulle quali si costruiscono gli algoritmi di ML.



Reti neurali multistrato (ML-NN)

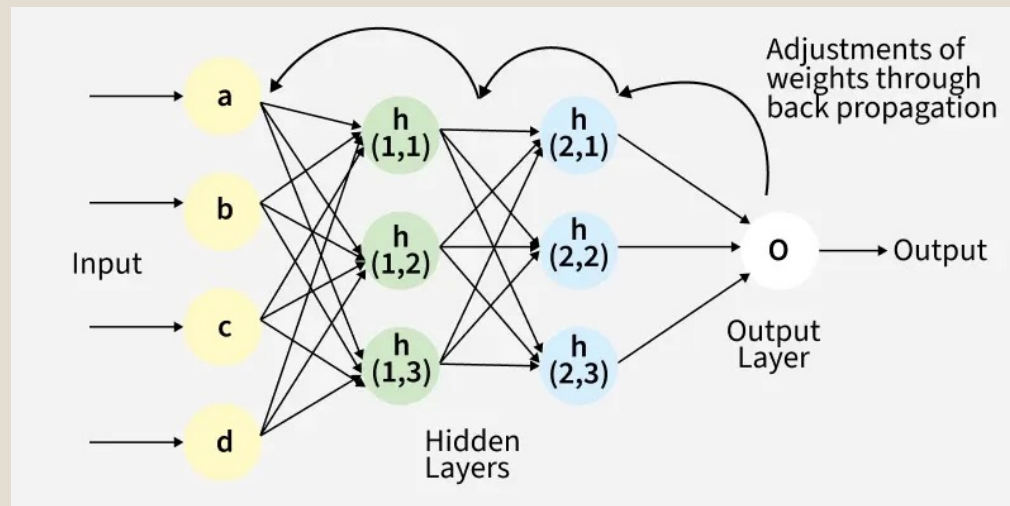
1. Una **rete** è un insieme di elementi comunque connessi tra loro
2. Una **rete neurale** è una rete di neuroni simulati (ovvero percettroni)
3. Una **rete neurale multistrato** è una rete neurale che impiega più strati per fornire un output. P.e. un rete neurale a due strati ha uno strato interno (o strato nascosto) e uno strato esterno (o di output). Similmente, il nostro cervello ha dei neuroni interni che svolgono delle operazioni interne (o parziali) e dei neuroni esterni (p.e. responsabili del movimento dei muscoli).
4. Una rete neurale multistrato è detta **profonda** (**deep network**) se vi sono più strati di unità nascoste (o interne).

Come nei percettroni, ogni unità riceve degli input pesati. Mentre però il percettrone scarica o no (0/1), una rete neurale calcola una somma che può o meno avvicinarsi a 1.



Backpropagation

- Mentre l'IA subsimbolica entrava in crisi, alcuni ricercatori ispirati da Rosenblatt continuarono le loro ricerche fino a dimostrare nei primi anni Ottanta che un algoritmo di apprendimento generale fosse possibile. Lo chiamarono **retropropagazione** (**backpropagation**)
- Si tratta di un algoritmo che misura l'errore in uscita e lo reinvia indietro chiedendo di modificare i pesi in modo da ridurlo a ogni passaggio. L'apprendimento consiste appunto in questo processo di progressiva riduzione dell'errore. In realtà, come faccia a farlo per mezzo delle unità nascoste non è affatto così chiaro (bisognerebbe valutare tutti i numerosi pesi a ogni passaggio!).



Anni '90

Schematicamente, dunque, vi sono due ambiti di sviluppo:

- quello **simbolico**, che ricalca la logica del cervello, che ha portato allo sviluppo di sistemi esperti, utili in ambiti basati sulla procedura;
- quello **subsimbolico**, basato sulle reti neurali, più adatto per compiti percettivi o motori (riconoscere una voce, afferrare una palla, comprendere una firma,...), di cui neanche noi conosciamo la procedura.

Sembrerebbe ovvio pensare a **sistemi ibridi** in grado di cogliere il meglio di entrambi: ed è effettivamente così che si è andati avanti negli anni Novanta e primi Duemila.

Ma i risultati, sebbene a volte interessanti, non sono stati rivoluzionari. Sempre più si sono imposti **metodi subsimbolici ispirati alla statistica e alla teoria della probabilità**, che hanno condotto allo sviluppo delle tecniche di **ML** oggi di successo.



Deep Blue

- Negli anni Novanta IBM si impegnò alla costruzione di un computer dedicato alle partite di scacchi.
- **Deep Blue** è stato il primo calcolatore a vincere una partita a scacchi contro un Campione del Mondo in carica, **Garry Kasparov**, con cadenza di tempo da torneo tra il 1996 e il 1997.

La forza di Deep Blue deriva principalmente dalla sua straordinaria potenza computazionale:

- un computer a parallelismo massivo a 30 nodi basato su RS/6000, supportato da 480 processori specifici VLSI progettati per il gioco degli scacchi. L'algoritmo per il gioco degli scacchi è scritto in linguaggio C e gira sotto un sistema operativo AIX: esso è capace di calcolare 200 milioni di posizioni al secondo.

Le sue funzioni di valutazione erano inizialmente scritte in forma generale, con molti parametri da definire. I valori ottimali per questi parametri furono poi determinati dal sistema stesso, analizzando migliaia di partite di campioni. Prima del secondo incontro, la conoscenza degli scacchi del programma era stata finemente migliorata dal gran maestro di scacchi Joel Benjamin.



Deep Blue

Polemiche di Kasparov

- Dopo il match perso, Kasparov disse che alcune volte gli era parso di notare nelle **mosse** della macchina intelligenza e creatività così profonde da non riuscire a comprenderle. Avanzò anche il **sospetto** che la macchina avesse avuto un **"aiuto" umano** durante la partita, sospetto che ritornò più volte, sia quando si seppe che la macchina non era posta nella stanza nella quale si disputava la partita ma ad alcuni chilometri di distanza e che quindi i dati venivano inviati da terzi, sia per il fatto che al campione russo non furono mai forniti i tabulati sull'attività del computer, che lui aveva richiesto secondo gli accordi della sfida.

- **Critica dell'IA**

Malgrado i grandi sforzi economici e intellettuali l'IBM non riuscì a convincere come tale macchina si potesse applicare ad altri campi del mondo intelligibile e quindi sprigionare la sua intelligenza generale.



Solo nel **2016** AlphaGo riuscì a battere il campione di **Go** !!



Un nuovo campo: il riconoscimento

Easy things are hard (Marvin Minsky)



Lo stesso Minsky aveva proposto il primo progetto di riconoscimento di oggetti (**Summer Vision Project**).

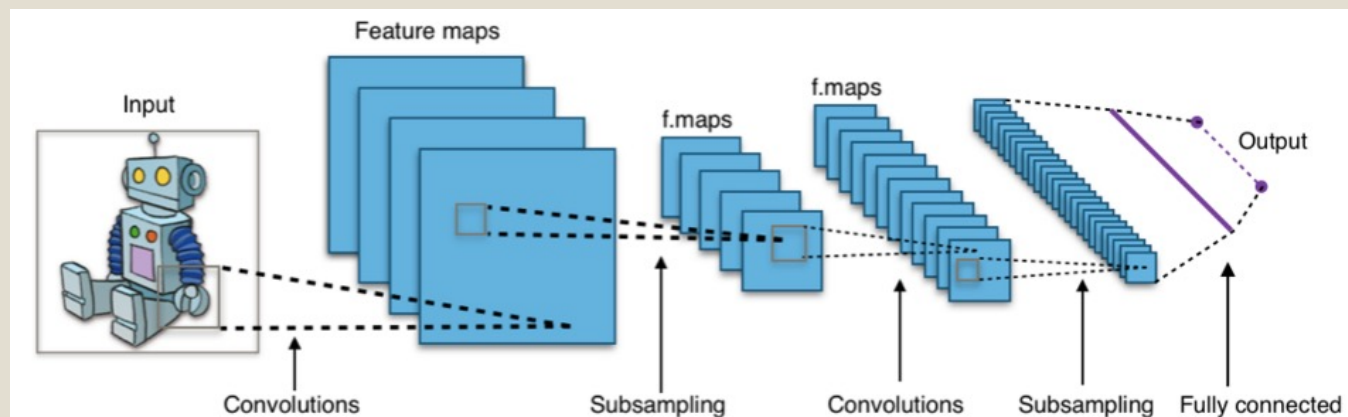
Il riconoscimento di oggetti, facile per l'uomo, sembra difficilissimo per un computer. Sebbene la computer vision sia progredita molto resta oggi ancora molto difficile distinguere tra un cane e un gatto con probabilità elevata.

Ma in tale settore sono arrivati dagli anni Duemila i migliori risultati, soprattutto grazie alla nota *triplice accelerazione*:

1. Disponibilità di grandi dati
2. Sviluppo di architetture di calcolo più veloci
3. Sviluppo delle Deep Neural Networks (DNN)

Deep Learning

- Con tale termine si intende l'uso di **Deep Neural Networks (DNN)** ovvero reti neurali con vari strati nascosti. Il fatto di lavorare con strati deep rende tali reti più simili al funzionamento del *cervello*.
- La prima DNN fu **Cognitron**, sviluppata da Kunihiro Fukushima. Ma fu con **Yann LeCun** che negli anni Ottanta si posero le basi per gli enormi sviluppi di oggi. Egli propose le **reti neurali convolutive (ConvNet)**, in analogia con il funzionamento del sistema visivo umano.
- Una ConvNet è un tipo di rete neurale artificiale feed-forward in cui il pattern di connettività tra i neuroni è ispirato dall'organizzazione della corteccia visiva animale: essa funziona con più strati di neuroni simulati (unità) collegati in una catena input-output. Ogni unità si attiva con un certo valore di attivazione a seconda dell'input e del peso.



ConvNet

- La rete è detta **convolutiva** poiché l'output è ottenuto dalla convoluzione dell'input (campo recettivo) per un peso (1: da luce a buio; 0: omogeneità; -1: da buio a luce), che migliora i contorni.
- L'uscita di tale convoluzione è un oggetto detto **mappa di attivazione**. Essa viene inviata in una rete neurale tradizionale (*modulo di classificazione*) che dà delle **percentuali di confidenza** delle categorie di oggetti conosciuti dalla rete (con che confidenza somiglia a un cane o a un gatto).
- È così che funziona il programma di ricerca immagini di Google: inserendo una qualsiasi nostra foto possiamo chiedere di riconoscerla. La confronterà con il suo archivio e dirà con quale immagine essa è più simile.



ImageNet

- Il business del «riconoscimento di immagini» diventò così fruttuoso che si organizzarono delle gare periodiche (*Visual Recognition*) cui i ricercatori sottopongono i loro algoritmi.
- Il gruppo di LeCun propose il database **ImageNet**, che sfruttava due recenti innovazioni:
 - Il riconoscimento per associazione di WordNet (proposto dalla giovane Fei-Fei Li)
 - L'aumento esponenziale di «immagini etichettate» grazie all'**Amazon Mechanical Turk**.



AlexNet

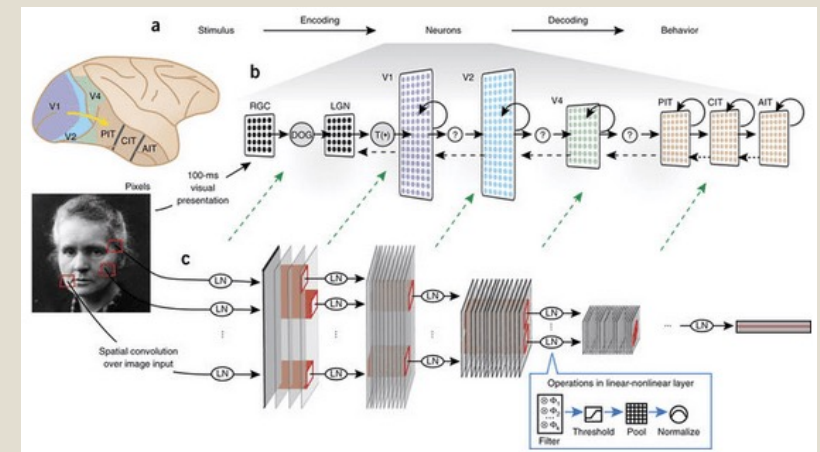
- La vera svolta si ebbe però con un nuovo algoritmo detto **AlexNet** (dal nome dell'inventore Alex Krizhevsky) il quale, contrariamente agli algoritmi precedenti, recuperava le DNN (che inizialmente sembravano non adatte al riconoscimento). AlexNet vinse la competizione nel 2012.

Da questo momento in poi e a tutt'oggi, il Deep Learning è il metodo di riferimento dell'IA.

Google entrò pesantemente in campo, acquisendo il gruppo di LeCun in blocco (allora alla NYU) e cominciò il predominio sul riconoscimento di immagini, tanto che oggi si può parlare di «**problema risolto**».

Ancora una volta, il successo è stato dovuto a una combinazione dei tre elementi chiave:

1. DNN
2. Big Data
3. GPU



Intervallo di confidenza: 90-95%

Superare le capacità umane?

- Se l'intervallo di errore delle ConvNet è ormai inferiore al 5% (limite umano), possiamo dire che «la macchina supera l'uomo»?

In realtà, le percentuali date si riferiscono infatti a una valutazione detta “*top five*”, ovvero il successo è valutato tale se almeno una su 5 delle categorie di categorizzazione è giusta.

Se volessimo un riconoscimento 1-su-1 il tasso di successo scenderebbe clamorosamente a poco più dell'80%!

Anche la stessa stima sull'uomo è puramente empirica e forse poco accurata basandosi su esperimenti di volontari....

Il tipo di errori tra le due classi (AI e uomo) sono inoltre molto diversi e spesso imparagonabili!

Mai fidarsi degli annunci dei media!!!



Limiti delle ConvNet

- La necessità dell'aiuto umano: **SUPERVISED LEARNING**
- Non è vero che una ConvNet è interamente artificiale, in quanto la **fase fondamentale di etichettatura viene ancora fatta da umani!** Tale fase genera dei c.d. *iperparametri*, cioè dei parametri scelti al di fuori della macchina, ovvero dall'uomo: sono questi che definiscono come debba avvenire l'apprendimento (numero di strati, grandezza dei campi, tasso di apprendimento, etichettatura,...). Tale fase di scelta è detta *tuning* o **ottimizzazione della rete neurale** e non può essere automatizzata. Si può anzi dire che tale fase “umana” abbia un che di artistico o addirittura di “cabalistico” (come l'ha definito il capo dell'IA di Microsoft).
- Il problema della coda lunga
 - Se cresce il numero degli elementi sulla coda d'errore il loro impatto può diventare trascurabile
 - Principali rischi:
 - AUTO A GUIDA AUTONOMA
 - HACKING



Overfitting



“foto con animale”

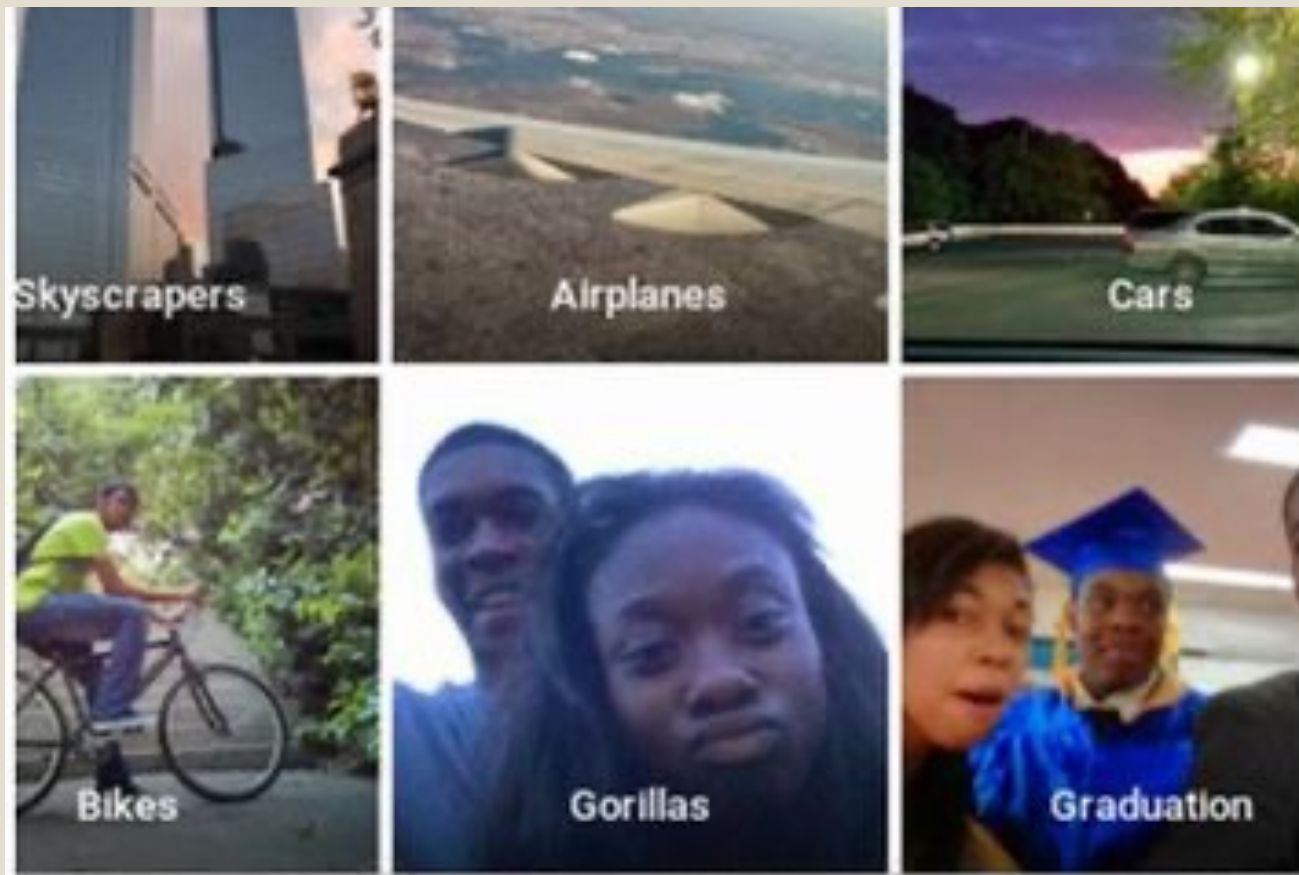


“foto senza animale”

Cosa ha veramente riconosciuto?

- L'animale
- Lo sfondo (in genere gli animali hanno uno sfondo sfocato)

Errori imbarazzanti



Source: Google Mistakenly Tags Black People as ‘Gorillas,’ Showing Limits of Algorithms, By [Alistair Barr](#) July 1, 2015, The Wall Street Journal.

Bias di addestramento

- L'addestramento viene tipicamente fatto su umani di razza caucasica (la maggior parte delle pubblicità e dei personaggi del web), causando imbarazzi e sviste gravi in un numero di casi non trascurabile. Se pensiamo che oggi con il riconoscimento facciale autorizziamo le nostre transazioni finanziarie o accediamo in aeroporto....!
- Altri pregiudizi imbarazzanti: classificare come donna chiunque fosse ritratto in una cucina con un grembiule da cuoco (pregiudizio!).
- Chi decide cosa sia un pregiudizio?

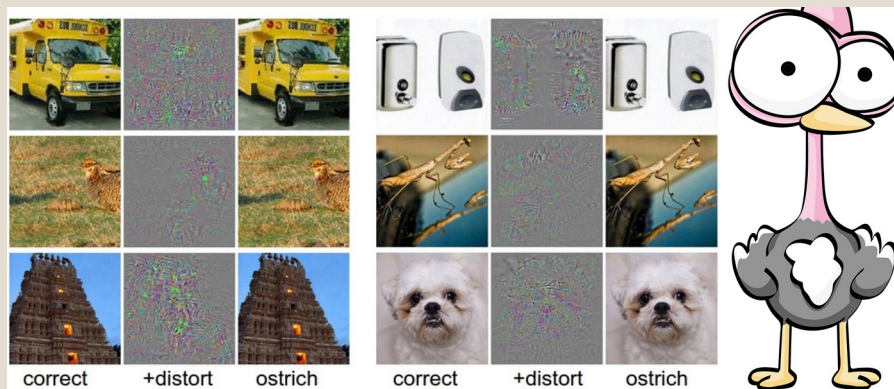


Ragazza

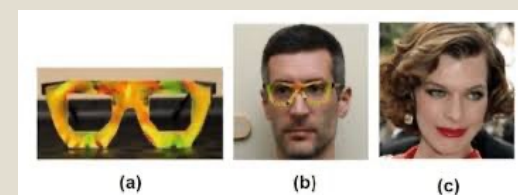


Persona in cucina

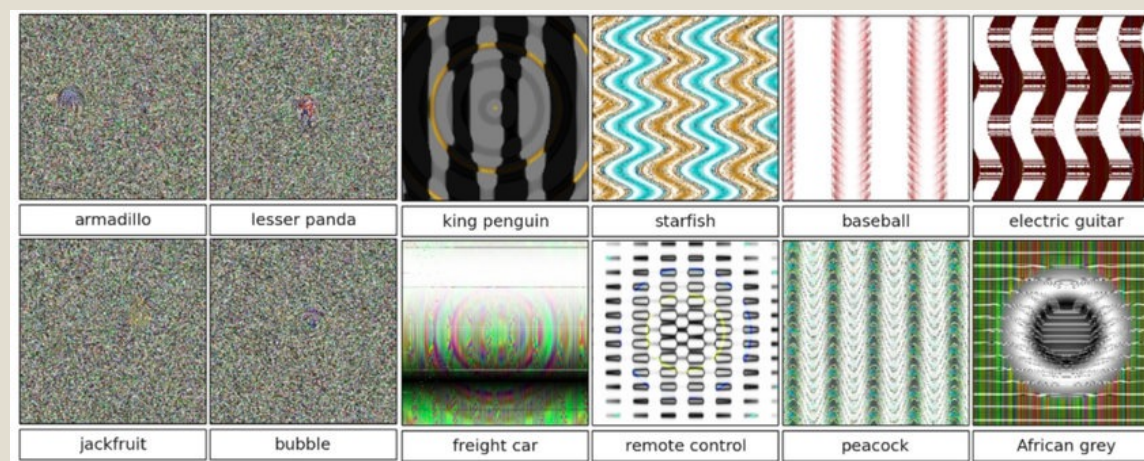
Ingannabilità: apprendimento avversario



Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., and Fergus, R. **Intriguing properties of neural networks**. ICLR (2013).



Occhiali ad hoc: l'uomo è identificato come l'attrice Milla Jovovich (!!)

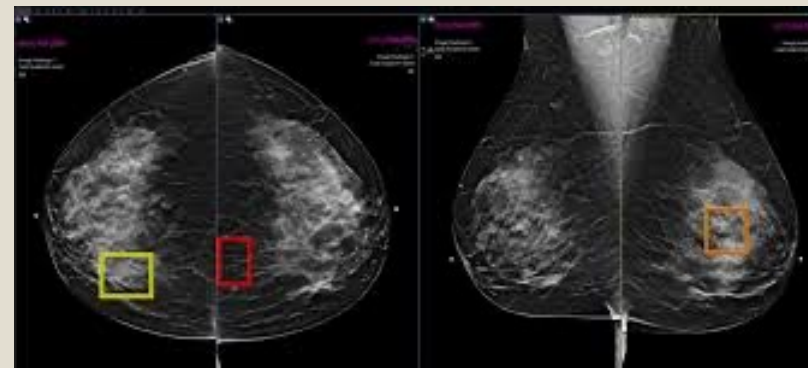


Nguyen A, Yosinski J, Clune J. Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images. In Computer Vision and Pattern Recognition (CVPR '15), IEEE, 2015.

Ingannabilità: rischi



Subtle perturbations cause a neural network to misclassify stop signs as speed limit 45 signs, and right turn signs as stop (source: <https://spectrum.ieee.org/slight-street-sign-modifications-can-fool-machine-learning-algorithms>)



Disease detection

**COSA CAPISCE
REALMENTE LA
MACCHINA ?**

Occorre avvicinarsi ancora
all'intelligenza umana....il
problema della *comprensione*

Giocare ma sul serio: il Reinforcement Learning

Per rendere l'apprendimento più simile a quello umano i programmatori usano le tecniche del *gaming*.

Una tecnica di apprendimento usata da sempre con umani e animali si chiama **condizionamento operante**: consiste nel “premiare i comportamenti che ci piacciono e ignorare quelli che non ci piacciono”. La sua applicazione al ML ha creato il cosiddetto apprendimento per rinforzo (*Reinforcement Learning*, RL). Diversamente dall'apprendimento supervisionato non ci sono etichette ma il programma di addestramento (detto agente) esegue azioni all'interno di un ambiente (p.e. una simulazione al computer). In seguito a qualche azione (non tutte!) riceve delle ricompense.



DeepMind

- L'esempio più notevole di uso del contesto game per sviluppare algoritmi AI è quello di **Demis Hassabis**, enfant prodige degli scacchi e poi del gaming, fondatore nel 2010 di **DeepMind**, poi vendita a Google nel 2014, quindi premio Nobel 2024 in Chimica. La sua idea era di realizzare un'integrazione tra “apprendimento per rinforzo” (in particolare Q-learning) e deep neural networks, per creare un sistema automatico per giocare con i videogiochi Atari.
- DeepMind produsse poi AlphaGo
- Gli algoritmi sviluppati per il gaming sono naturalmente fondamentali per raccogliere pubblicità e finanziamenti: ma gli stessi boss dell'IA sanno che il vero successo deve prima o poi esportarsi ad attività di interesse più generale per diventare rivoluzionari.

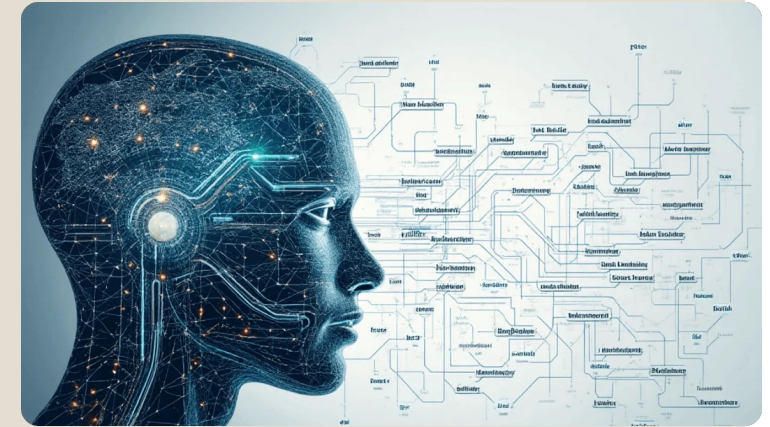


IA e linguaggio naturale

Uno dei problemi più brillantemente risolti di recente nell'ambito dell'IA è quello dell'*elaborazione del linguaggio naturale* (**NLP, Natural Language Processing**), alla base degli attuali chatbot e di tutte le applicazioni che concernono:

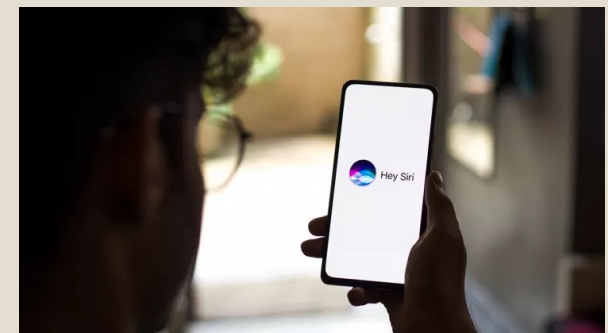
- Riconoscimento vocale
- Ricerca web
- Risposta automatica a domande (chatbot)
- Traduzione automatica

2012: articolo congiunto dei team di ricerca di Microsoft, Google, IBM e Università di Toronto sul riconoscimento vocale (*Geoffrey Hinton* et al., Deep Neural Networks for Acoustic Modeling in Speech Recognition, IEEE SIGNAL PROCESSING MAGAZINE, November 2012).



RICONOSCIMENTO VOCALE

ANDROID



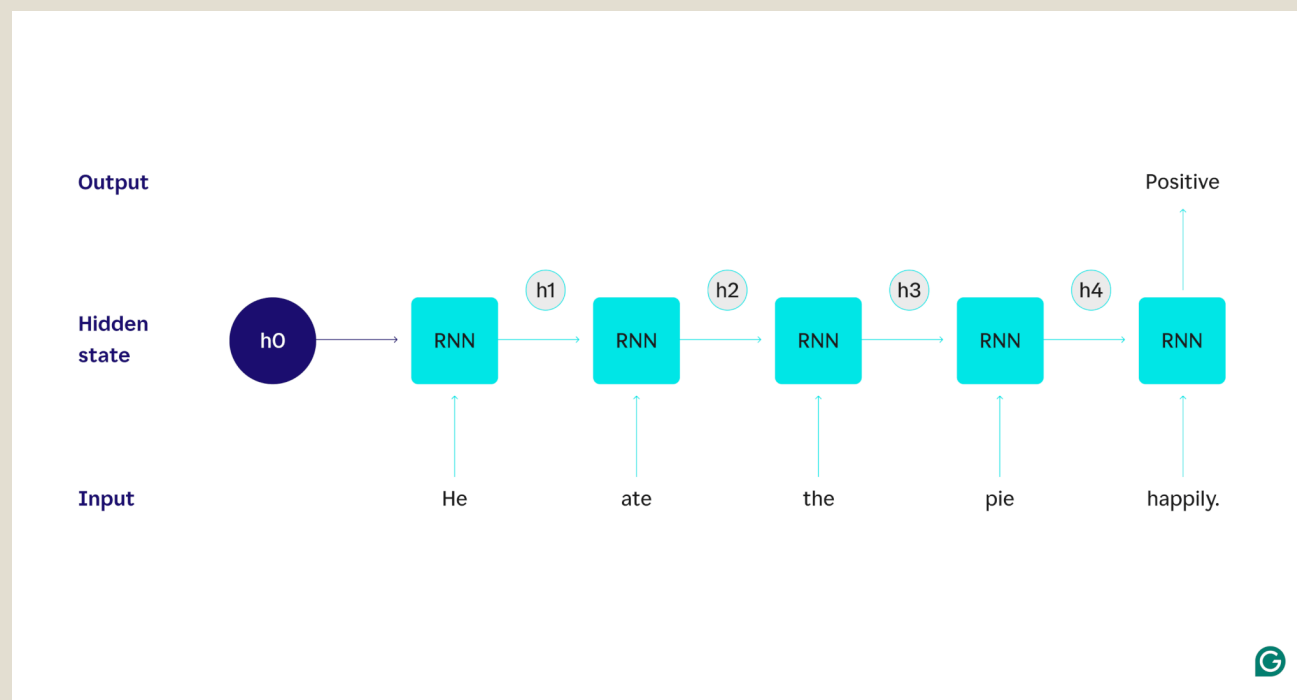
RICONOSCIMENTO VOCALE

APPLE

IA e linguaggio naturale

L'addestramento per il linguaggio richiede tecniche diverse da quelle delle immagini. Anzitutto le immagini hanno un numero fisso di pixel, mentre per definizione le frasi hanno lunghezza imprevedibile.

Le tecniche si basano sulle **RNN (*Recurrent Neural Networks*)**, ispirate alle *teorie sull'interpretazione delle sequenze* da parte del cervello delle neuroscienze.

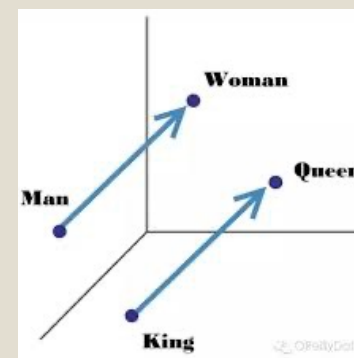


A differenza delle tradizionali RN, la RNN calcola la propria attivazione sia sull'input sia sull'attivazione delle unità nascoste al passo temporale precedente: ciò permette una sorta di *ricordo* di quanto appena letto, così che due parole consecutive possano legarsi tra loro “semanticamente”.

IA e linguaggio naturale

Google nel 2013 propone l'algoritmo chiamato *word2vec* (*word to vector*): l'idea è di associare le parole (*token*) a dei numeri, anzi a dei **vettori** (**embedding**) in uno *spazio semantico*

```
1 nel      [1,0,0,0,0,0,0,0,0,0,0,0,0,0]
2 mezzo    [0,1,0,0,0,0,0,0,0,0,0,0,0,0]
3 del      [0,0,1,0,0,0,0,0,0,0,0,0,0,0]
4 cammin   [0,0,0,1,0,0,0,0,0,0,0,0,0,0]
5 di       [0,0,0,0,1,0,0,0,0,0,0,0,0,0]
6 nostra    [0,0,0,0,0,1,0,0,0,0,0,0,0,0]
7 vita     [0,0,0,0,0,0,1,0,0,0,0,0,0,0]
8 mi       [0,0,0,0,0,0,0,1,0,0,0,0,0,0]
9 ritrovai [0,0,0,0,0,0,0,0,1,0,0,0,0,0]
10 per     [0,0,0,0,0,0,0,0,0,1,0,0,0,0]
11 una     [0,0,0,0,0,0,0,0,0,0,1,0,0,0]
12 selva   [0,0,0,0,0,0,0,0,0,0,0,1,0,0]
13 oscura  [0,0,0,0,0,0,0,0,0,0,0,0,1,0]
```



Gli LLM moderni, basati sull'architettura **Transformer**, utilizzano tecniche di embedding più avanzate e sofisticate: un **embedding contestuale** anziché un embedding fisso.

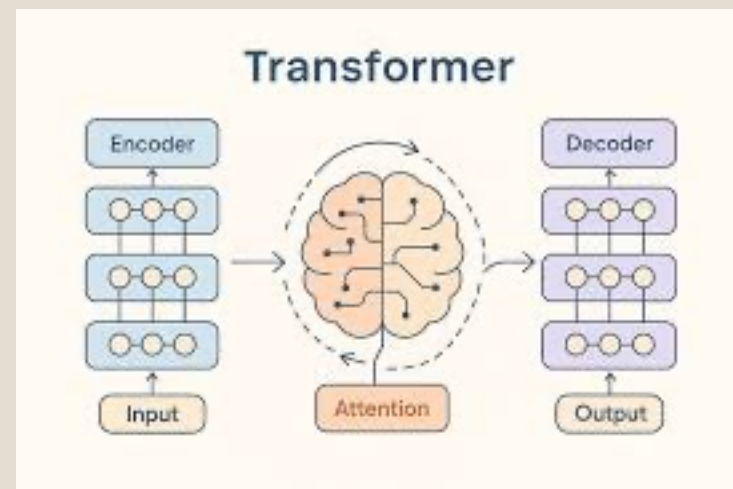
token corrente	finestra di contesto
nel	(nel,mezzo) (nel,del)
mezzo	(nel, mezzo) (mezzo,del) (mezzo,cammin)
cammin	(mezzo, cammin)(del,cammin)(cammin,di)(cammin,nostra)
di	(del,di)(cammin,di)(di,nostra)(di,vita)
nostra	(cammin,nostra)(di,nostra)(nostra,vita)
vita	(di,vita)(nostra,vita)

Transformer

Nel **2017**, l'articolo *"Attention is All You Need"* di **Vaswani et al.** ha introdotto il **Transformer**, un'architettura che ha rivoluzionato il campo dell'elaborazione del linguaggio naturale (NLP). I Transformer hanno superato i limiti delle RNN e LSTM permettendo l'elaborazione in parallelo delle sequenze di dati e migliorando significativamente le prestazioni su vari compiti di NLP.

Sulla base del Transformer sono stati sviluppati i **modelli linguistici di grandi dimensioni (Large Language Models, LLM)** che prevedono un **token** o una sequenza di token, a volte molti paragrafi di token previsti. Un **token** può essere una parola, una subword (un sottoinsieme di una parola) o anche un singolo carattere. Gli LLM fanno previsioni migliori rispetto ai modelli linguistici N-grammi o alle reti neurali ricorrenti perché:

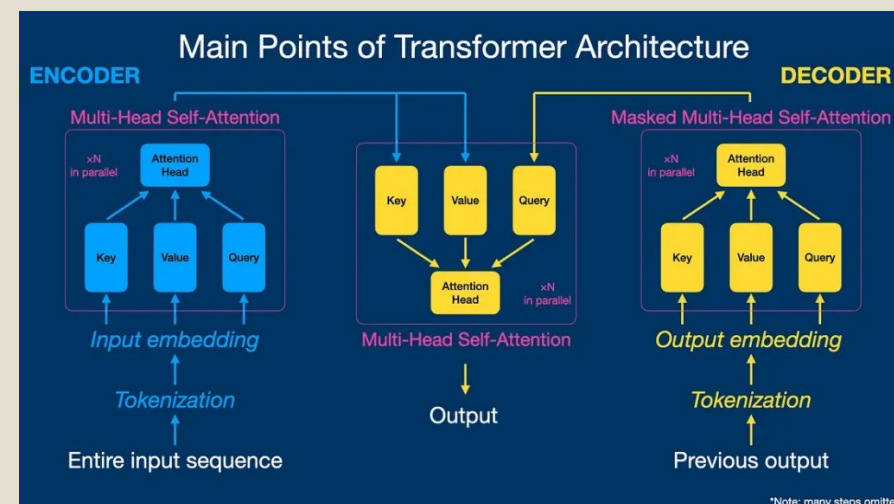
- Gli LLM contengono molti più **parametri** rispetto ai modelli ricorrenti.
- Gli LLM raccolgono molti più contesti.



Che cosa è un Transformer

Un **Transformer** è un'architettura di rete neurale rivoluzionaria che elabora i dati sequenziali (come il testo) in parallelo, anziché una parola alla volta, grazie a un meccanismo chiamato **auto-attenzione** (*self-attention*), permettendo una **comprensione profonda del contesto e delle relazioni** tra gli elementi della sequenza per compiti complessi come traduzione, generazione di testo e analisi, superando i limiti dei modelli precedenti come le RNN.

- Elaborazione parallela: A differenza delle vecchie RNN che leggevano parola per parola, i Transformer vedono l'intera sequenza **contemporaneamente**, identificando quali parole sono più importanti l'una per l'altra.
- Meccanismo di attenzione (*Self-Attention*): Il cuore del Transformer, assegna pesi diversi a ciascuna parola in base alla sua rilevanza nel contesto, permettendo al modello di "prestare attenzione" alle parti giuste della frase.
- Architettura Encoder-Decoder: Spesso è composto da un Encoder (che trasforma l'input in una rappresentazione numerica) e un Decoder (che genera l'output da tale rappresentazione).
- Efficienza: L'elaborazione parallela lo rende più veloce ed efficiente, specialmente per sequenze lunghe.

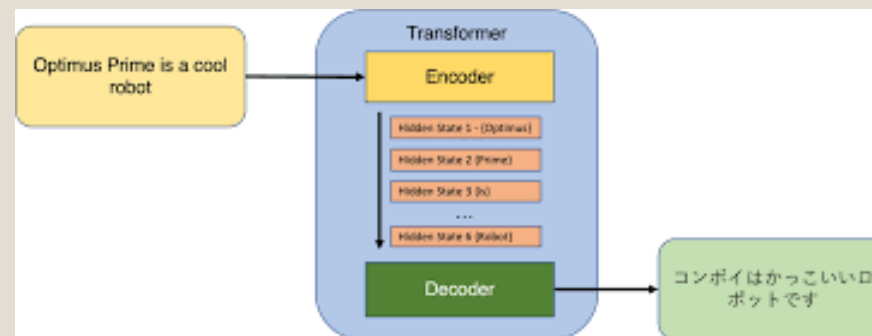


Usi del Transformer

Ha rivoluzionato l'NLP (Natural Language Processing) dal 2017, anno del suo introduzione ("Attention is All You Need").

Applicazioni principali

- **Traduzione automatica:** Comprendere il contesto globale per traduzioni più accurate.
- **Generazione di testo:** Creare contenuti coerenti e di alta qualità (es. GPT).
- Analisi del sentimento, riconoscimento vocale, sintesi di proteine e molti altri compiti legati alla sequenza.



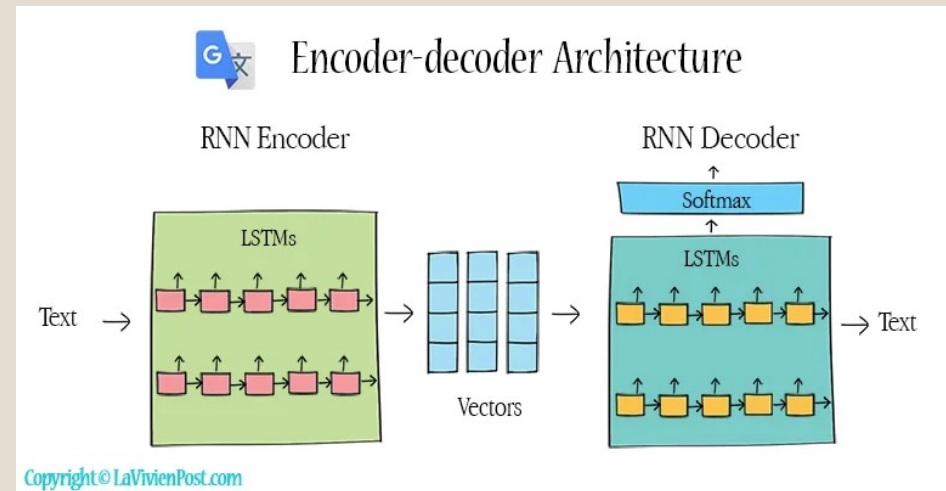
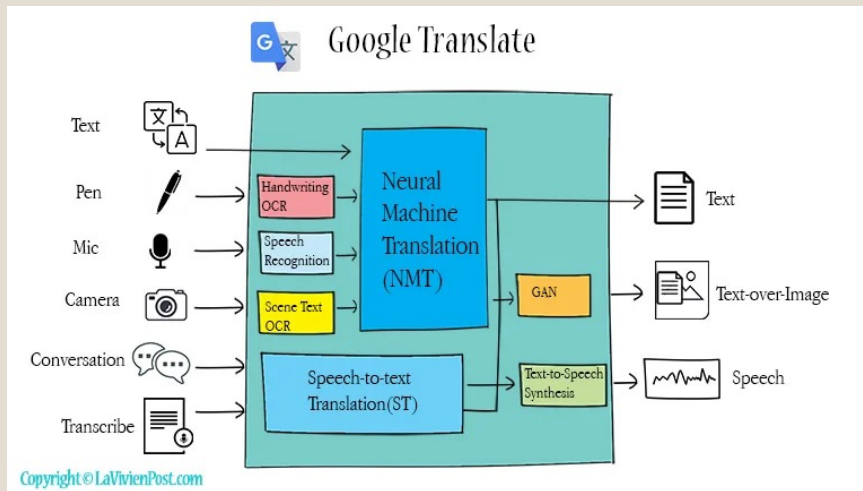
Vantaggi chiave

Comprensione del contesto: Cattura relazioni a lungo termine che i modelli precedenti perdevano.

Parallelizzazione: Calcoli più veloci grazie all'elaborazione simultanea.

Google Translate: cosa c'è sotto il cofano?

Usa una tecnica basata sul Deep Learning detta *traduzione automatica neurale* (Google *Neural Machine Translation*, *GNMT*), poi utilizzata ovunque: il Transformer traduce un'intera frase alla volta e garantisce una maggiore precisione del contesto.



Dal 2026, il servizio integra tutte le tecnologie derivate dai LLM per offrire traduzioni più naturali e gestire contesti complessi o idiomi. Oltre al testo, queste tecniche neurali alimentano funzioni avanzate come la **traduzione visiva istantanea** tramite fotocamera e la **traduzione simultanea** della voce in tempo reale.

Chatbot



Un chatbot è una **combinazione di interfaccia, linguaggio naturale, logica decisionale e dati**, con livelli crescenti di intelligenza in base alle tecnologie utilizzate:

Elaborazione del Linguaggio Naturale (NLP): Permette al chatbot di capire il significato delle parole e delle frasi umane, anche con errori o variazioni.

Machine Learning (ML): I chatbot più avanzati usano il ML per imparare dalle interazioni passate, migliorando la precisione e la pertinenza delle risposte nel tempo.

Risposte basate su regole/script (per i più semplici): Seguono flussi predefiniti per gestire domande frequenti (FAQ) e compiti basilari.



IA generativa: nuova fase storica

L'IA generativa (**Intelligenza Artificiale generativa**) è una branca dell'IA che crea nuovi contenuti originali:

- **testo**
- **immagini**
- **musica**
- **codice**



Lo fa analizzando **enormi set di dati** e apprendendo schemi per produrre output simili ma inediti, spesso indistinguibili da quelli umani, usando modelli come GPT, DALL-E o Midjourney, rivoluzionando settori creativi, tecnologici e aziendali, pur sollevando questioni etiche su accuratezza e diritti d'autore.



Una nuova fase storica grazie a:

1. **Big Data**
2. **Nuove architetture di calcolo**

Big Data

I **Big Data** sono enormi volumi di dati, strutturati e non, generati a velocità elevatissime da persone e macchine, che richiedono tecnologie avanzate (come IA e ML) per essere raccolti, elaborati e analizzati, estraendo informazioni preziose per prendere decisioni strategiche, ottimizzare processi, personalizzare offerte e innovare il business, trasformando dati grezzi in valore competitivo in vari settori.



Caratteristiche principali (5Vs)

- **Volume:** Enormi quantità di dati (terabyte, petabyte).
- **Velocity:** Flusso continuo e rapido di dati in tempo reale.
- **Variety:** Dati strutturati (database), semi-strutturati (XML) e non strutturati (testo, immagini, video, social media).
- **Veracity:** Qualità e affidabilità dei dati
- **Value:** L'importanza di estrarre insight utili.

Potenza computazionale

Una **GPU (Graphics Processing Unit)** è un circuito elettronico specializzato progettato originariamente per velocizzare la creazione di immagini e video attraverso un processore progettato per accelerare la creazione di immagini in un frame buffer, destinato all'output su un dispositivo di visualizzazione..

Le GPU sono fondamentali per l'IA perché la loro architettura a calcolo parallelo accelera enormemente l'addestramento e l'esecuzione dei modelli di apprendimento profondo (deep learning), gestendo grandi quantità di dati contemporaneamente.

Le GPU permettono di sviluppare rapidamente applicazioni di IA, dal riconoscimento vocale alla guida autonoma, fino alla creazione di contenuti generativi, grazie alla loro **potenza di elaborazione, scalabilità e flessibilità**, anche se stanno emergendo hardware più specifici (**acceleratori AI**) e soluzioni ottimizzate per l'**edge computing**.



Perché le GPU sono cruciali per l'AI

- Parallelismo massivo: Le GPU sono progettate per eseguire migliaia di operazioni in parallelo, un'architettura perfetta per le operazioni matematiche ripetitive su grandi set di dati (tensori) tipiche delle reti neurali.
- Accelerazione computazionale: Riducono enormemente i tempi di addestramento dei modelli di machine learning, permettendo iterazioni più veloci nello sviluppo dell'AI.
- Addestramento e inferenza: Sono essenziali sia per la fase di "imparare" (addestramento) sia per quella di "utilizzare" (inferenza) i modelli AI.
- Specializzazione hardware: I produttori stanno integrando core specifici (es. Tensor Core di NVIDIA) per accelerare le operazioni di intelligenza artificiale.
- Efficienza energetica: Rispetto alle CPU, possono essere più efficienti per i carichi di lavoro AI paralleli.



Applicazioni principali nell'AI

AI generativa: Creazione di immagini, video, testo (es. modelli linguistici di grandi dimensioni - LLM).

Visione artificiale: Riconoscimento facciale, miglioramento delle immagini (super-risoluzione, upscaling).

Analisi dati e simulazioni: Gestione di Big Data, simulazioni scientifiche e analisi predittive.

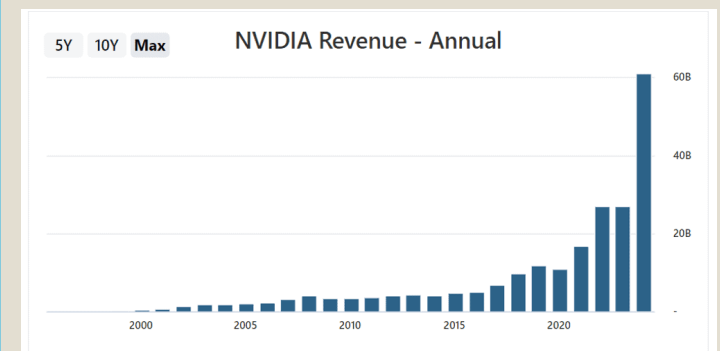
Elaborazione del linguaggio naturale (NLP): Traduzione, analisi del sentiment, chatbot.

Principali attori e tecnologie

NVIDIA: Leader di mercato, con architetture come A100, RTX e tecnologie CUDA.

AMD e Intel: Stanno sviluppando soluzioni competitive per l'AI.

TPU (Tensor Processing Unit): Sviluppate da Google, sono chip ancora più specializzati per il deep learning, ma le GPU restano la scelta più diffusa e versatile.



IA e società

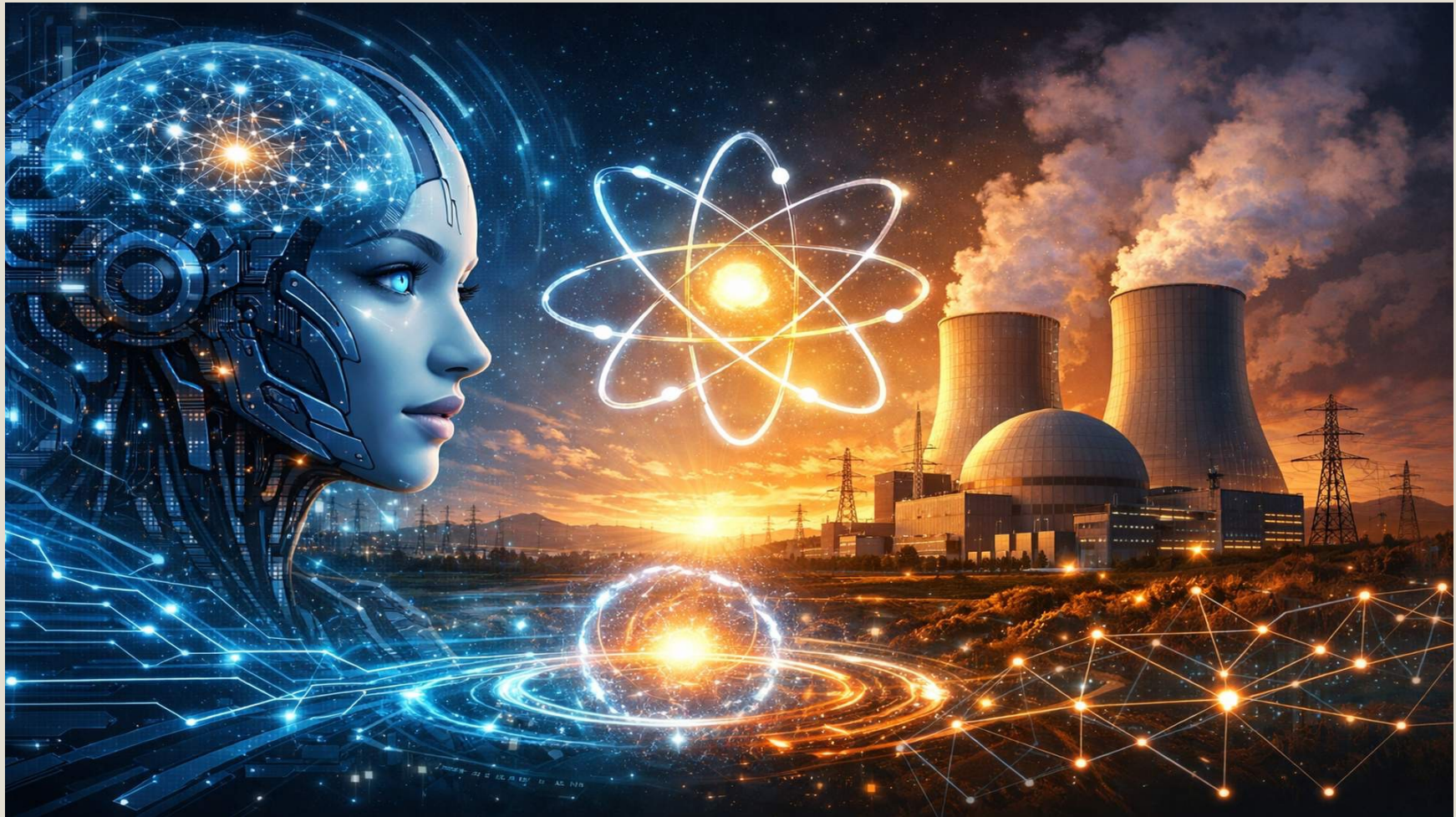
L'IA sta rivoluzionando la società trasformando **aziende, lavoro, istruzione e vita quotidiana**, offrendo efficienza e innovazione ma sollevando questioni etiche e sociali cruciali, che l'Europa affronta con normative come l'**AI Act** per garantire uno sviluppo sicuro, responsabile e trasparente, bilanciando benefici e rischi e promuovendo la alfabetizzazione digitale dei cittadini per un'integrazione consapevole.

Sfide Etiche e Sociali

- **Privacy e Dati:** Gestione responsabile e protezione dei dati personali.
- **Preconcetti (Bias) ed Equità:** Evitare discriminazioni algoritmiche, garantendo inclusività.
- **Impatto sul Lavoro:** Gestione della transizione occupazionale e necessità di riqualificazione.
- **Regolamentazione:** Necessità di quadri normativi (come l'AI Act) per guidare lo sviluppo e l'adozione dell'IA in modo etico e legale.



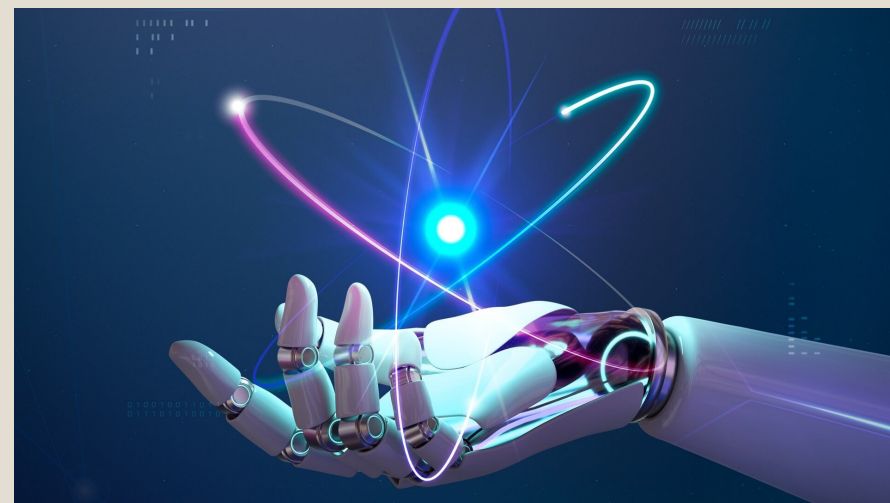
Applicazioni nel mondo nucleare



Cosa può fare l'IA per il Nucleare?

DOMANDE PRINCIPALI

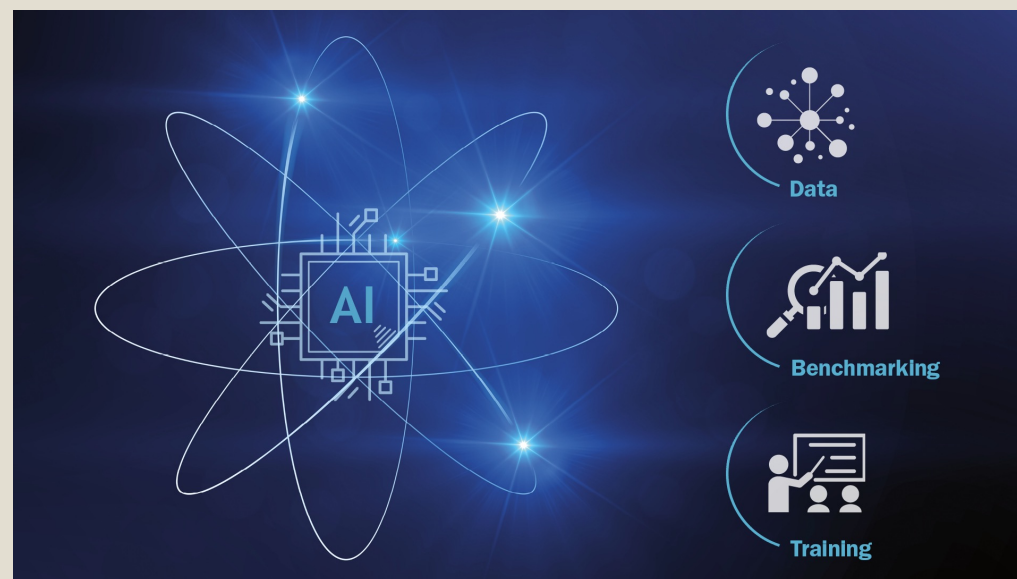
1. Perché è **necessaria** l'IA in campo nucleare?
2. In che modo affronterà o risolverà il problema di interesse e in che modo si **differenzia** dalle altre tecnologie?
3. Quali capacità dell'IA la rendono una soluzione **migliore** rispetto ad altre tecnologie?
4. Quali **azioni aggiuntive** devono essere intraprese per sviluppare, distribuire e implementare una tecnologia di IA?
5. Quali misure saranno richieste in termini **regolatori**?



Joint Project on AI Platform for Nuclear Research and Education (AIxpertise): 2026-28



Il progetto congiunto **AIxpertise** mira a costruire una **piattaforma strutturata e collaborativa** che consenta agli esperti di sfruttare l'Intelligenza Artificiale (IA) per il settore dell'energia nucleare con solidità e trasparenza. I partecipanti al progetto saranno coinvolti in attività pratiche su tre pilastri: **dati**, **benchmarking** e **formazione**. Amplieranno le conoscenze, raccoglieranno e gestiranno i dati e valuteranno gli strumenti per applicare soluzioni di IA in modo efficace e sicuro per il futuro del settore dell'energia nucleare.



Il progetto creerà una piattaforma collaborativa per riunire stakeholder provenienti da enti di ricerca, autorità di sicurezza, organizzazioni di supporto tecnico (TSO), industria, mondo accademico e aziende tecnologiche.

US-NRC and AI

L'NRC si impegna a livello esterno attraverso **collaborazioni internazionali, ricerca applicata, valutazioni della prontezza normativa e forum pubblici** per lo scambio tecnico. Le attività riflettono un **approccio coordinato** alla comprensione delle tecnologie di intelligenza artificiale emergenti, alla valutazione delle loro implicazioni per la regolamentazione nucleare e al dialogo con la più ampia comunità tecnica e normativa, in linea con la continua evoluzione di questi strumenti.

CONSIDERATIONS FOR DEVELOPING ARTIFICIAL INTELLIGENCE SYSTEMS IN NUCLEAR APPLICATIONS

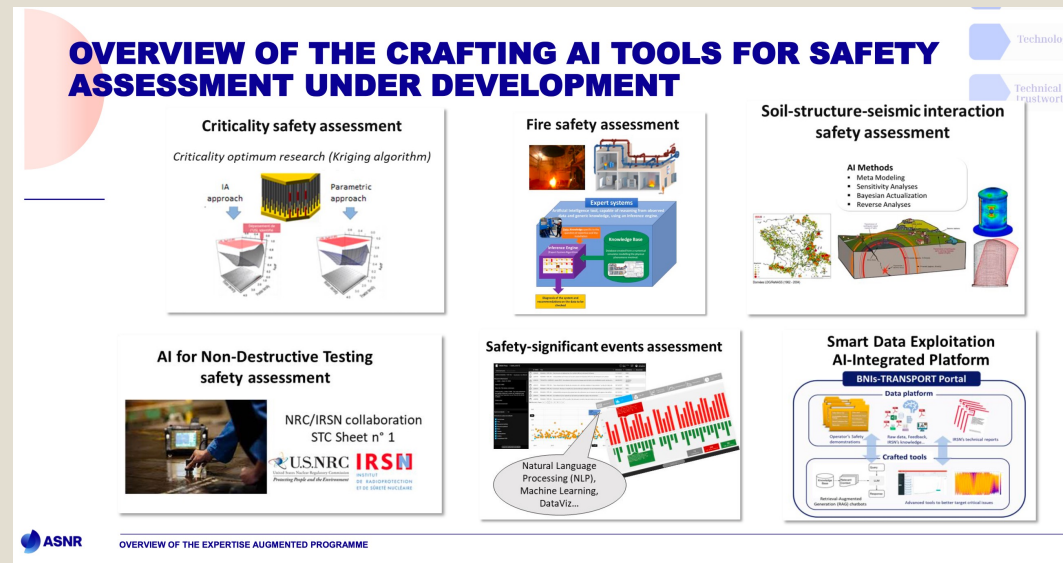
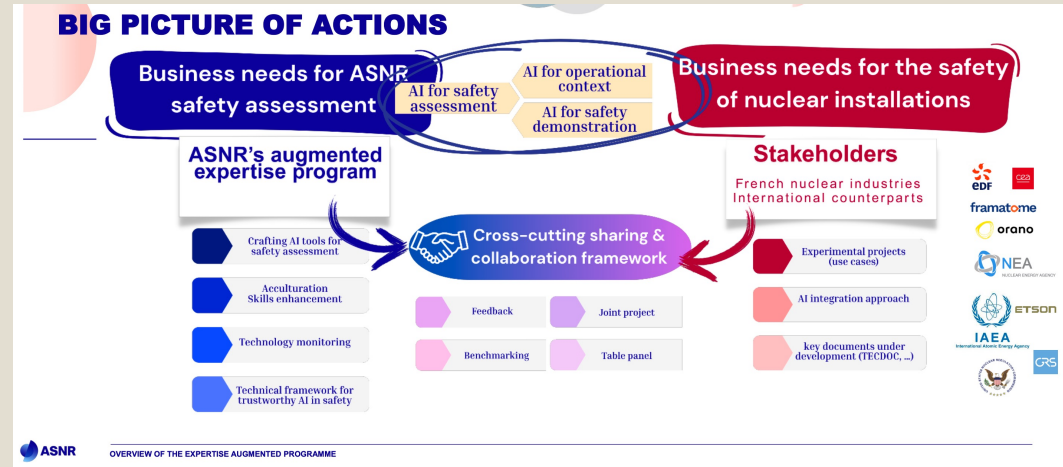
SEPTEMBER 2024

Canadian Nuclear Safety Commission
UK Office for Nuclear Regulation
US Nuclear Regulatory Commission



ASNR ADDRESSING THE TECHNICAL CHALLENGES OF AI IN SUPPORT OF NUCLEAR SAFETY ASSESSMENT (2025)

- Considerazioni per **l'implementazione** di applicazioni di intelligenza artificiale nel settore dell'energia nucleare
- **Impatto** dell'intelligenza artificiale sulla **sicurezza**
- Gruppo di lavoro ISOP sull'intelligenza artificiale (**ISOP AI WG**)
- Analisi congiunta dei vantaggi delle **piattaforme dati** e intelligenza artificiale, con un prossimo white paper congiunto da redigere con l'NRC e il GRS degli Stati Uniti.



IAEA Guidelines 2025

Obiettivi

- **Panoramica** delle diverse applicazioni di intelligenza artificiale e apprendimento automatico basate sui dati e sulla fisica nel settore dell'energia nucleare che hanno il potenziale per migliorare il livello di autonomia quando implementate in una centrale nucleare;
- **Identificazione** fattori tecnici, umani, stakeholder, normativi e di altro tipo (inclusi hardware, software e infrastruttura) per consentire una più ampia implementazione di soluzioni di intelligenza artificiale e l'accettazione da parte degli utenti finali;
- Importanza dell'**integrità dei dati**;
- **Coinvolgimento** degli stakeholder e la valutazione del rischio durante tutto il processo di sviluppo e implementazione dell'intelligenza artificiale.

Considerations for Deploying Artificial Intelligence Applications in the Nuclear Power Industry

IAEA Guidelines 2025



Possibili benefici

- Aumentare l'automazione operativa e il processo decisionale ausiliario, riducendo il carico di lavoro degli operatori delle centrali nucleari;
- Rafforzare la capacità di monitoraggio e analisi dello stato delle apparecchiature, dei sistemi e delle unità, migliorando la sicurezza delle centrali nucleari;
- Fornire strumenti più efficienti e flessibili al personale addetto alla gestione e all'esercizio delle centrali nucleari.

Main areas of AI capability

Shallow models (e.g. decision trees, linear regression, support vector machines, Bayesian models)

Deep learning networks

Natural language processing

Teaching and learning

Monitoring and diagnostics

Text analytics using large language models

Assistance to non-destructive evaluation inspections

Fault diagnosis and predictive maintenance

Operational and decision support

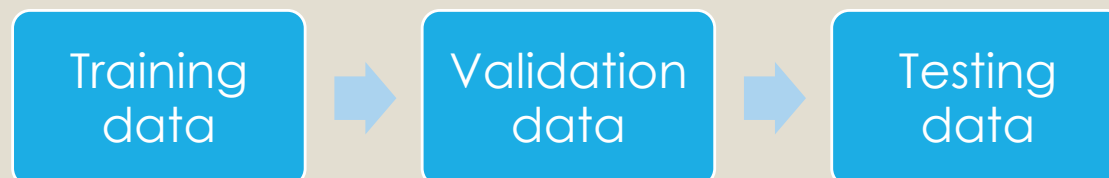
Human performance optimization tools for operators

Sensor state on-line monitoring

Nuclear safeguards applications

Operator actions, reliability analysis and probabilistic risk assessment

IAEA Guidelines 2025

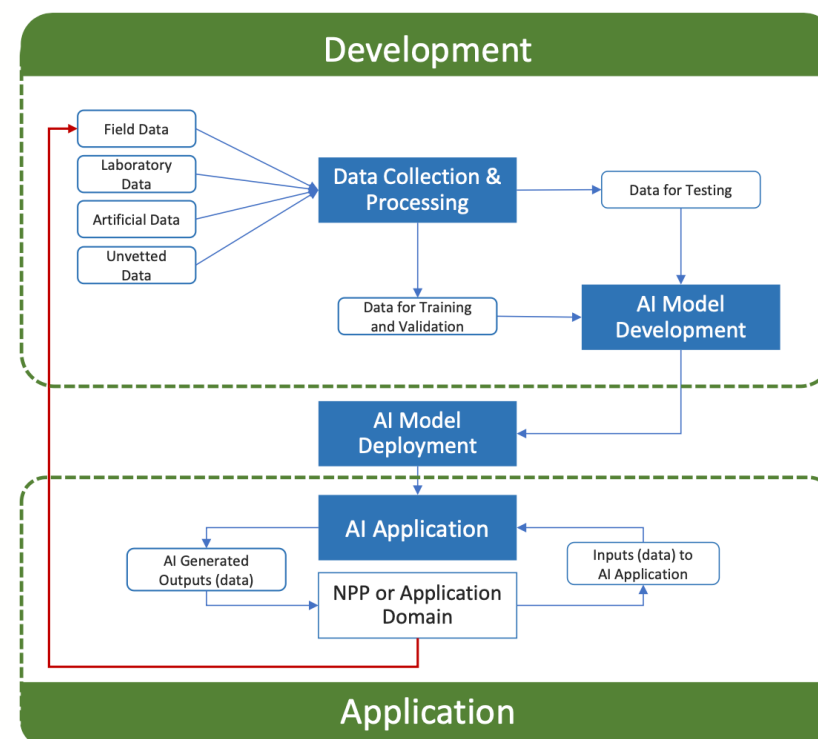


Gestione dei dati

I **dati** svolgono un ruolo fondamentale nelle applicazioni di intelligenza artificiale.

Le applicazioni di intelligenza artificiale dipendono dai dati per apprendere e funzionare, e il loro successo dipende in larga misura dal grado di accuratezza, completezza, coerenza e pertinenza dei dati correlati. Indipendentemente dalla qualità di un algoritmo di intelligenza artificiale, i suoi risultati possono rivelarsi inaffidabili se si utilizzano dati inadeguati.

Data usage through the life cycle of AI applications



IAEA Guidelines 2025



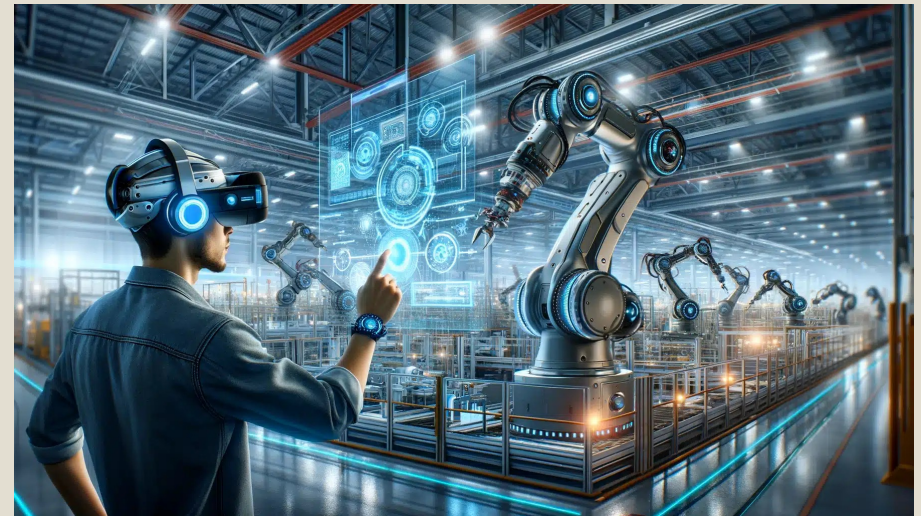
In sintesi dunque le linee-guida IAEA sottolineano che :

- Necessaria comunicazione chiara tra gli stakeholder su vantaggi e limiti dell'uso dell'**IA nel settore nucleare**.
- Necessario **coinvolgimento di esperti, operatori e regolatori**, insieme a un'implementazione graduale, cruciale per costruire fiducia e affidabilità.
- Il **ciclo di vita dell'IA** richiede team interdisciplinari, integrazione con sistemi esistenti e un livello di formalizzazione proporzionato ai **rischi di sicurezza, operativi ed economici**.
- Prima dell'adozione è necessaria una **valutazione tecnica, economica e dei rischi**, seguendo un approccio pianificato e metodico.
- La **qualità e gestione dei dati** sono determinanti per il successo dell'IA; è essenziale l'allineamento tra dati di addestramento e lo scenario applicativo reale.
- La **condivisione dei dati** favorisce lo sviluppo di soluzioni efficaci.
- I regolatori stanno sviluppando **approcci basati sul rischio**; per l'operatività servono «spiegabilità», formazione del personale e infrastrutture adeguate.
- Nonostante i progressi, **permangono sfide e opportunità** per un impiego sicuro ed efficace dell'IA nell'industria nucleare.

IA e Nucleare: punti chiave

Novità dell'IA: pro e contro

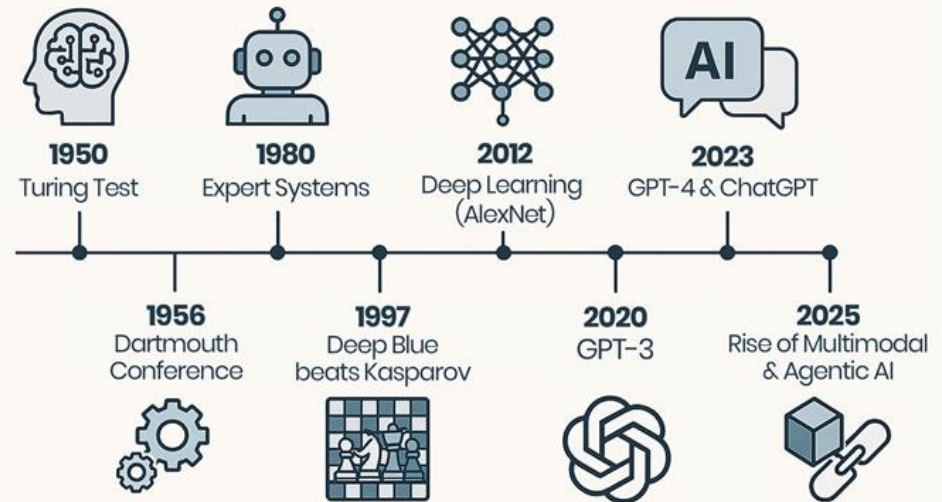
- **Dalle regole ai dati:** non programmiamo più la macchina, **addestriamo** la macchina.
- **Potenza vs intelligenza:** il salto è dovuto soprattutto a GPU e parallelizzazione (Transformer), non necessariamente a una maggiore “**comprensione**” dell'intelligenza.
- **Il problema della scatola nera:** cruciale per il nucleare.
 - Come fidarsi di un modello «non spiegabile»?
 - Come licenziare sistemi «inverificabili»?



- no al controllo automatico *black-box* di funzioni di safety;
- no a decisioni autonome senza supervisione umana (*human-in-the-loop*).

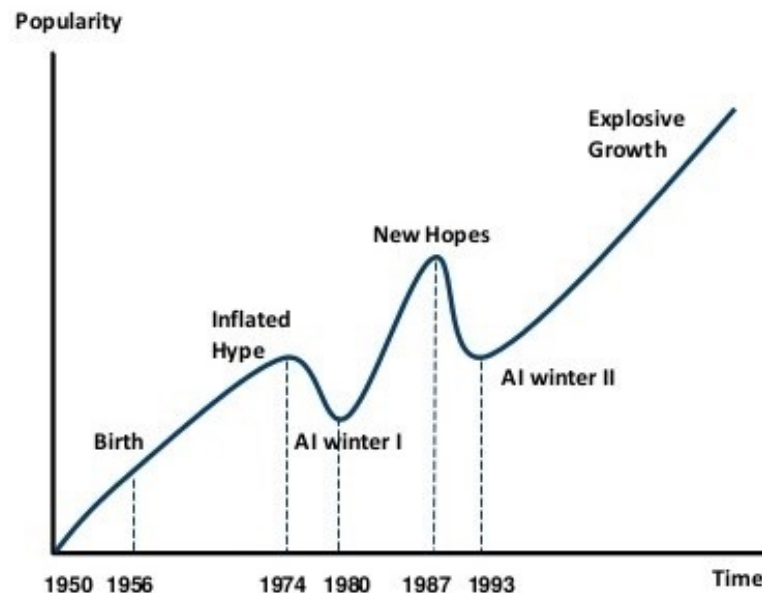
Riassumendo

HISTORY OF ARTIFICIAL INTELLIGENCE



Storia ciclica: summers and winters

AI HAS A LONG HISTORY OF BEING "THE NEXT BIG THING" ...



Timeline of AI Development

- **1950s-1960s**: First AI boom - the age of reasoning, prototype AI developed
- **1970s**: AI winter I
- **1980s-1990s**: Second AI boom: the age of Knowledge representation (appearance of expert systems capable of reproducing human decision-making)
- **1990s**: AI winter II
- **1997**: Deep Blue beats Gary Kasparov
- **2006**: University of Toronto develops Deep Learning
- **2011**: IBM's Watson won Jeopardy
- **2016**: Go software based on Deep Learning beats world's champions



Domande aperte

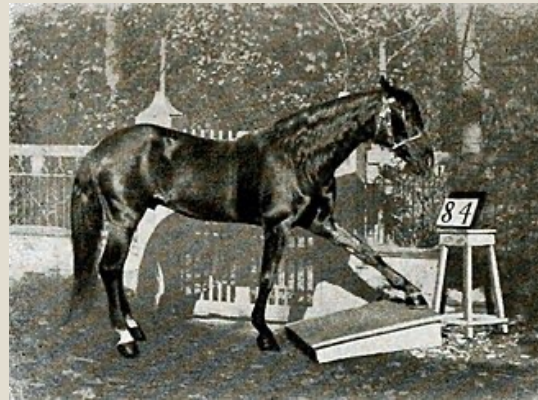
- A quando la guida autonoma? Sì ma a che livello?
- Disoccupazione di massa?
- Creatività o meccanicità?
- A quando un'IA Generale?
- La civiltà dell'IA farà scomparire la civiltà umana?
- Dovremo aver paura della potenza delle macchine o della loro vulnerabilità?
- Dobbiamo essere soddisfatti del livello oggi raggiunto? *Sì, ma sostanzialmente **TUTTI** i problemi posti nel meeting di Dartmouth del 1955 sono insoluti!*

Un progetto di lungo termine

- Progetto culturale → coinvolge tutti gli umani («*la nuova elettricità?*»)
- Futuro: uomo e società → etica (Leggi di Asimov) e regolamentazione
- Cambio di nome?
- Tre scenari futuri



The Matrix



L'astuto Hans
(*Der kluge Hans*)



L'uomo e la macchina
Woody Allen, *Il dormiglione*, 1973



Take-home messages

- l'IA attuale è un potente **interpolatore statistico**, non un'entità pensante;
- nel nucleare la safety richiede modelli **physics-informed**, non solo correlazioni statistiche;
- il futuro è l'**intelligenza aumentata**: l'uomo resta il decisore, l'IA è lo strumento che filtra il rumore dai dati.

Bibliografia minima

- S. Russell, P. Norvig, “Artificial Intelligence: a Modern Approach,” Prentice Hall, 1995 (<http://www.cs.berkeley.edu/~russell/aima.html>)
- Nick Bostrom, Superintelligenza, Bollati Boringhieri, 2018
- Melanie Mitchell, L’Intelligenza artificiale, Einaudi 2022
- Luciano Floridi, Etica dell’intelligenza Artificiale, Raffaello Cortina Editore, 2022.
- Anil Anathaswamy, Why Machines Learn, Penguin, 2025

Disclaimer

Le immagini pubblicate in questa presentazione sono di proprietà dei rispettivi autori e/o sono state reperite da risorse online con l'intento di essere utilizzate per fini personali, informativi o didattici. Non vi è pertanto alcun uso commerciale.