

Oratorio Don Bosco (Gardolo) Lunedì 11 novembre 2024

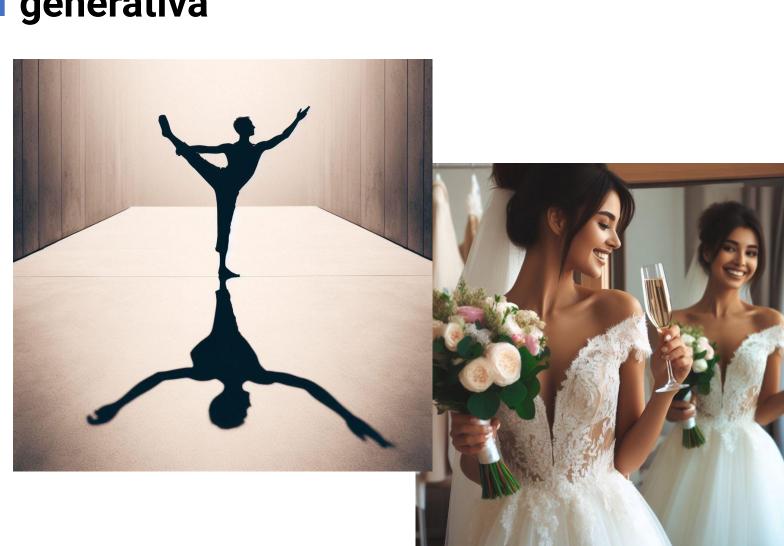
Intelligenza Artificiale: Per un approccio critico e consapevole

Gianluca Schiavo

Center for Augmented Intelligence Fondazione Bruno Kessler

Valeria Fabretti

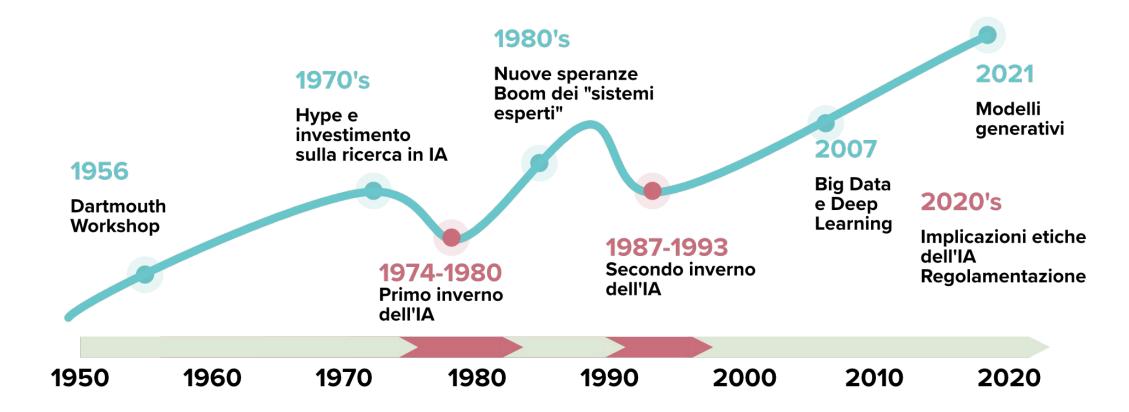
Center for Religious Studies Fondazione Bruno Kessler







ai evoluzione e fattori





ai definizione

L'Intelligenza Artificiale è un insieme di **tecnologie** e processi che possono, per un dato insieme di obiettivi **definiti da agenti umani**, fare previsioni, raccomandazioni o decisioni che influenzano ambienti **reali o virtuali**. I sistemi di intelligenza artificiale **interagiscono** con noi e agiscono sul nostro ambiente, direttamente o indirettamente.

Spesso possono dimostrare dei gradi di **autonomia** e possono **adattare** il loro comportamento imparando a conoscere il contesto in cui operano.









Ruolo delle persone



Interazione reale o virtuale



Diversi gradi di autonomia

John McCarthy



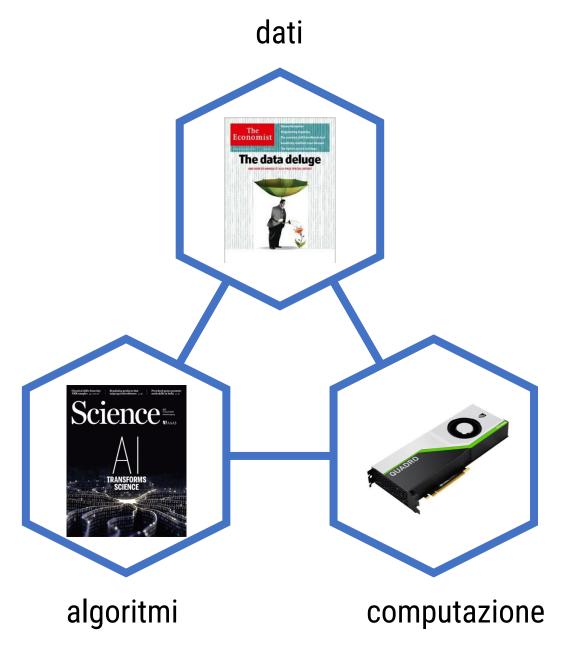
A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence

August 31, 1955

John McCarthy, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon



ai evoluzione e fattori





big data strutturati e non





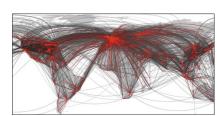


328 Eb





90%



1M

totale dati prodotti fino al 31/12/2023

dati generati quotidianamente nel 2024

dei dati esistenti sono stati prodotti negli ultimi 2 anni

ore di streaming video caricate ogni minuto nel 2023

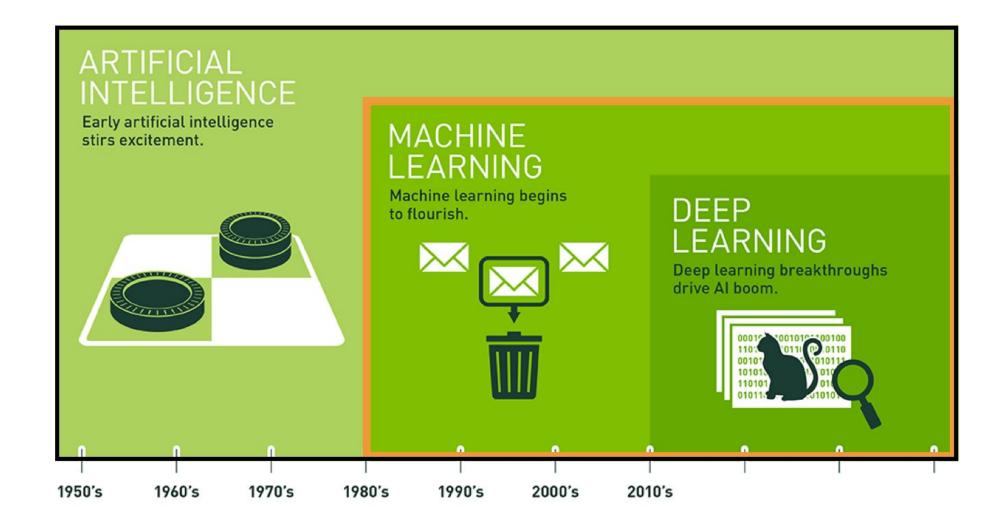




apprendimento algoritmi



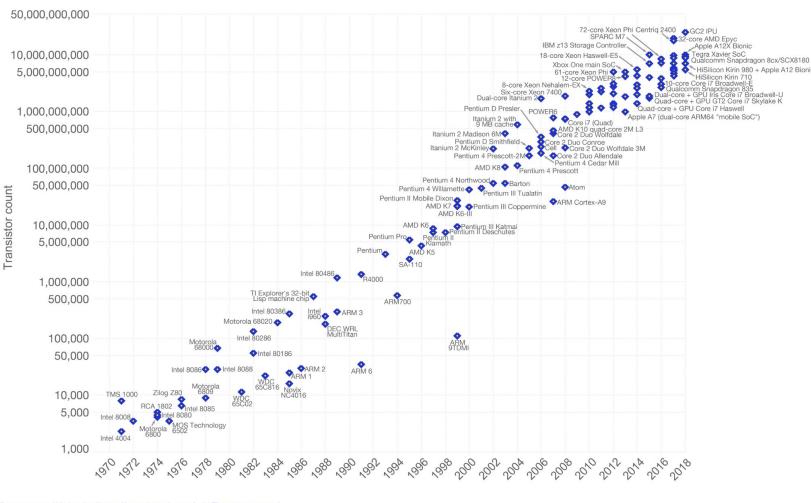
L'arte di sviluppare **algoritmi** e tecniche che permettano ai computers di **apprendere**, ovvero di acquisire conoscenza costruendo **modelli** dei dati osservati attraverso **l'inferenza** dagli esempi con l'obiettivo di utilizzarli poi in predizione.





ai potenza computazionale







| ai | complessità



capacità di apprendimento...

per compiti complessi (e critici!) come la guida autonoma, i sistemi di ai apprendono a riconoscere pedoni, veicoli, segnali, elementi delle strade e a anticipare azioni future.



...guidata dai dati...

ogni componente è addestrata per prova ed errore da esempi: migliaia di scene, in cui gli elementi sono stati etichettati; milioni di simulazioni di possibili eventi, fino ad ottimizzare il sistema.



...verso la predittività

dati nuovi, mai visti in addestramento, sono usati come test virtuale per valutare il sistema, per garantire accurate capacità di previsione sui dati futuri







ai the black box issue

Harvard Journal of Law & Technology Volume 31, Number 2 Spring 2018

THE ARTIFICIAL INTELLIGENCE BLACK BOX AND THE FAILURE OF INTENT AND CAUSATION

Yavar Bathaee*



molto spesso, i modelli di IA forniscono ottime performance, ma risulta quasi impossibile capire **perché funzionino**

01

Non c'è modo per determinare come il modello arrivi a una decisione o previsione 02

Non c'è modo di determinare quali informazioni sono **determinanti** per l'output del modello 03

Non c'è modo di ottenere una classifica delle variabili elaborate dal modello nell'ordine della loro importanza



Che cos'è l'IA Generativa?

L'Intelligenza Artificiale generativa (IA generativa o GenAI) si riferisce a modelli di Intelligenza Artificiale che hanno la capacità di creare contenuti.

Questi sistemi non si limitano a elaborare o analizzare informazioni, ma possono effettivamente generare nuovi dati o contenuti che non esistevano prima.



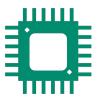








Tra le principali architetture utilizzate nelle tecnologie generative ci sono le **Generative Adversarial Networks** (GANs) e le **Reti neurali ricorrenti** (RNN) e i modelli **Transformer** (come GPT)



Tecnologie



Ruolo delle persone



Interazione reale o virtuale



Diversi gradi di autonomia



2019

Prime immagini di gatto generate da modelli IA (Generative adversarial networks)



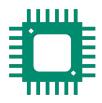


2022

thiscatdoesnotexist.com Sito web che utilizza algoritmi di IA per creare immagini di gatti che non esistono nella realtà



L'Al generativa necessita di enormi set di dati per l'addestramento (ad esempio l'intero web), e spesso questi dati devono essere preparati, etichettati o revisionati da **persone**. Inoltre l'**intervento umano** è usato per ottimizzare le architetture neurali e i parametri per migliorare la qualità dei contenuti generati.



Tecnologie



Ruolo delle persone



Interazione reale o virtuale



Diversi gradi di autonomia



l'Al generativa può essere utilizzata in ambienti, personaggi e narrazioni virtuali, ma anche influenzare ambienti reali (ad esempio se integrata in robot o macchine industriali)





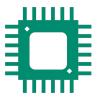


Ruolo delle persone





l'Al generativa può adattarsi a diversi contesti e richieste sulla base delle indicazioni fornite dall'utente ma anche da processi di **fine-tuning**



Tecnologie



Ruolo delle persone



Interazione reale o virtuale



Diversi gradi di autonomia



| <mark>ai</mark> | limiti



Bias

tendenza del modello a produrre risultati distorti o discriminatori, influenzati dalle caratteristiche dei dati di training



Disinformazione

le tecnologie di Al generativa possono essere utilizzate per diffondere disinformazione in modo rapido ed efficiente



"Allucinazioni"

il modello produce output immaginari o non realistici, spesso dovuti a limitazioni nel training o a input ambigui



Privacy e sicurezza

possibilità che i dati sensibili vengano utilizzati in modo improprio o che vengano generati contenuti falsi che possono danneggiare la privacy degli individui



ai limiti



Bias

tendenza del modello a produrre risultati distorti o discriminatori, influenzati dalle caratteristiche dei dati di training



Disinformazione

le tecnologie di Al generativa possono essere utilizzate per diffondere disinformazione in modo rapido ed efficiente



"Allucinazioni"

il modello produce output immaginari o non realistici, spesso dovuti a limitazioni nel training o a input ambigui



Privacy e sicurezza

possibilità che i dati sensibili vengano utilizzati in modo improprio o che vengano generati contenuti falsi che possono danneggiare la privacy degli individui



Gli stereotipi di genere: qualche anno fa ...



Edizioni Locali V Servizi V CORRIERE DELLA SERA ® ABBONATI A Accedi

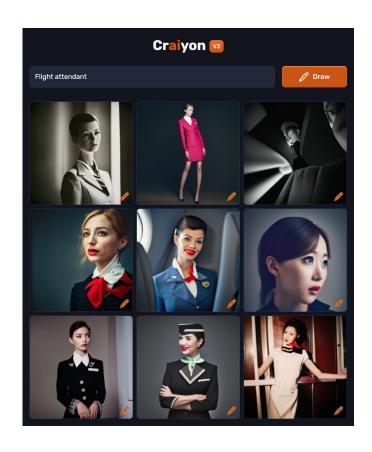
Il Natale «sessista» di ChatGPT: «Al papà regala una bottiglia di vino, alla mamma un set di pentole»

di Federico Cella

Abbiamo messo alla prova l'algoritmo generativo di OpenAl sui regali da fare ai parenti. E i suggerimenti dell'intelligenza artificiale lasciano a bocca aperta: tristi stereotipi. «È lo specchio della nostra società»







Gli stereotipi di genere nell'IA generativa





Le immagini create possono rappresentare degli stereotipi dovuti a bias inclusi nei dati di training







I sistemi di intelligenza artificiale generativa tendono a pregiudizi, stereotipi e riduzionismo anche quando si tratta di identità nazionali:

https://restofworld.org/2023/ai-image-stereotypes/

ai limiti



Bias

tendenza del modello a produrre risultati distorti o discriminatori, influenzati dalle caratteristiche dei dati di training



Disinformazione

le tecnologie di Al generativa possono essere utilizzate per diffondere disinformazione in modo rapido ed efficiente



"Allucinazioni"

il modello produce output immaginari o non realistici, spesso dovuti a limitazioni nel training o a input ambigui



Privacy e sicurezza

possibilità che i dati sensibili vengano utilizzati in modo improprio o che vengano generati contenuti falsi che possono danneggiare la privacy degli individui



Information disorder: la quantità delle informazioni aumenta continuamente e circola attraverso molte e diverse fonti, anche in modo automatico e incontrollato (difficile distinguere se creazioni umane)

Offerta di informazioni non richieste sulla base della profilazione

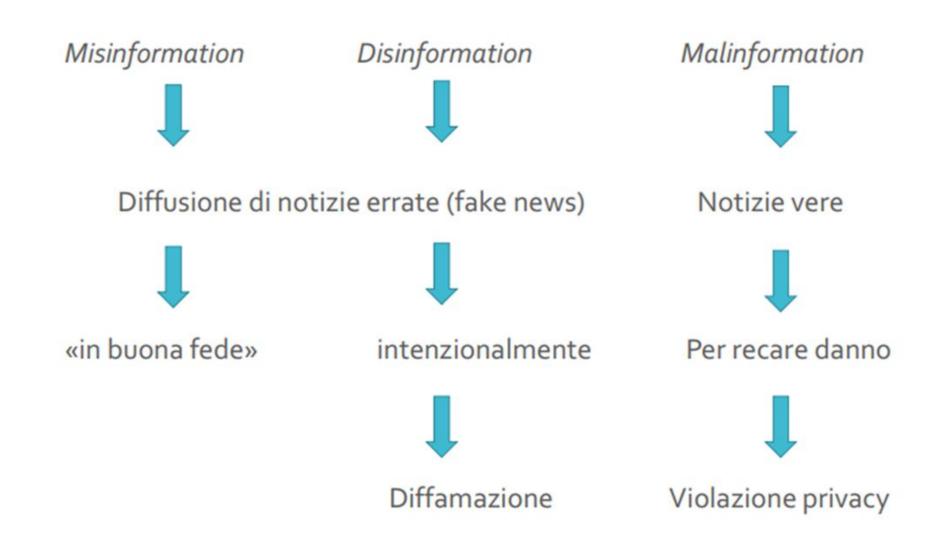
Creazione di «camere dell'eco» (Echo Chambers): ovvero ambienti virtuali all'interno di una piattaforma di condivisione online in cui l'attività di un utente si correla con quella dei suoi contatti.



Echo Chambers: come funzionano?

- 1. Abbiamo la tendenza ad aggregarci a persone con le stesse attitudini e interessi
- 2. Gli algoritmi rinforzano questa attitudine e, facendoci vedere messaggi, notizie e commenti verso i quali abbiamo già in precedenza mostrato interesse, ci conducono in spazi dove la nostra voce (ovvero i fatti e le opinioni di cui siamo già convinti/e) risuona incontrastata (come in una grotta dove i suoni riverberano)
- 3. Il numero di persone coinvolte in un processo virale è strettamente collegato alla dimensione della echo-chamber di cui fa parte







TikTok usato come motore di ricerca? "Rischio fake-news altissimo"

Rapporto di NewsGuard su 540 risultati di TikTok basati sulla revisione dei primi 20 risultati di 27 ricerche su argomenti di attualità. L'agenzia ha trovato che 105 video, il 19,4%, contenevano affermazioni false o fuorvianti

Fake e deep-fake

INTERNET DOMENICA 31 LUGLIO 2022

Il finto Tom Cruise di TikTok

Per molti anni, Miles Fisher soffrì la somiglianza con uno degli attori più celebri di Hollywood: finché non decise di sfruttarla



I'm taking lessons.



Dance-off X !



Never too old to learn!











Dopo lo stupore per la falsa immagine di Bergoglio, è importante chiedersi chi sia l'autore. Perché la sua storia - come quella di un ragazzo italiano che ha provato a imitarlo - ci aiuta a capire molte cose sulla nuova intelligenza artificiale generativa

29 MARZO 2023 AGGIORNATO ALLE 13:32



Una storia di fake ... o di creatività?



Perché ci crediamo?

Dinamiche note alla psicologia:

- È più facile credere ad una informazione che dubitare di essa (*Truth Bias*)
- Tendiamo a credere più facilmente a informazioni che troviamo ripetute spesso (Illusion of Truth)
- Possiamo voler trovare nelle notizie false delle conferme ai nostri stereotipi e pregiudizi (Confirmation Bias)
- Guardiamo con scetticismo le informazioni che ritraggono in modo negativo membri dei gruppi cui apparteniamo

Una questione individuale? Attenzione agli **impatti sociali:** la manipolazione dell'informazione può portare a scontri e **polarizzazioni**



Hate Speech

Una definizione (Consiglio d'Europa, Raccomandazione n. 15/2015 della commissione contro il razzismo e l'intolleranza)

- Fomentare, promuovere o incoraggiare, sotto qualsiasi forma, la denigrazione, l'odio o la diffamazione nei confronti di una persona o di un gruppo
- Sottoporre a soprusi, insulti, stereotipi negativi, stigmatizzazione o minacce una persona o un gruppo
- Giustificare tutte queste forme o espressioni di odio sulla base della "razza", del colore della pelle, dell'ascendenza, dell'origine nazionale o etnica, dell'età, dell'handicap, della lingua, della religione o delle convinzioni, del sesso, del genere, dell'identità di genere, dell'orientamento sessuale e di altre caratteristiche o stato personale

I discorsi d'odio sono diventati uno dei maggiori problemi da gestire su ogni tipo di piattaforma online. I gestori di piattaforme online adottano policy e provvedimenti per limitare questo fenomeno

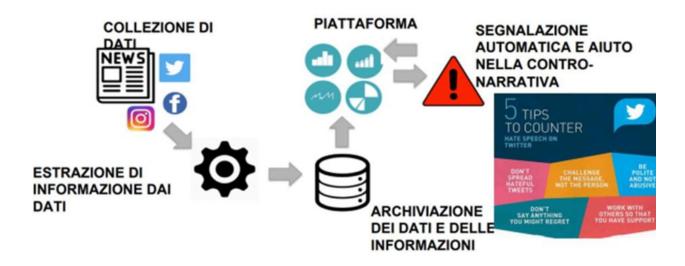
Dati, approfondimenti e materiali : https://www.retecontrolodio.org/





Il ruolo dell'IA nella propagazione e nel contrasto







Conoscere il mondo attraverso «i dati»?



L'IA si nutre di dati

Il dato non esiste in realtà, è una **costruzione** che tenta di indicare una parte di realtà in modo **probabilistico** (computazionale)

Anche moltiplicando i dati da cui l'IA può attingere e rendendoli sempre più sofisticati mancherebbe sempre una parte di **realtà** (es. sotto-rappresentazione delle minoranze): probabile ≠ vero

La conoscenza del mondo richiede invece relazione con- e interpretazione (senso e significato) dei contesti e delle situazioni: serve la nostra soggettività!

Un oggetto è reale solo alla fine di un processo complesso di ricerca (J. Dewey)



Concludendo

L'IA non esiste indipendentemente dagli esseri umani. È un campo in costante evoluzione, il cui sviluppo e impatto non è ancora chiaro

In quanto innovazione, è sempre guidata da valori che dipendono da chi la progetta e per quali scopi

Lo sviluppo e l'utilizzo di applicazioni di intelligenza artificiale comporta decisioni e competenze non solo "tecniche": saper immaginare le ricadute sociali e etiche (non cosa potremmo ma cosa dovremmo fare o non fare con l'IA)



Concludendo

Per una IA affidabile

- Intervento e sorveglianza umani
- Trasparenza
- Diversità, non discriminazione ed equità
- Benessere sociale e ambientale
- Riservatezza e governance dei dati
- Robustezza tecnica e sicurezza
- Accountability (responsabilizzazione)



Concludendo

Oltre le **visioni opposte**: lA come male assoluto vs lA come promessa di sviluppo Verso la **relazione critica e consapevole**: preservare **l'umano** nell'interazione con l'IA Alcune domande aperte (che non possiamo non farci)

- Quale idea dell'umano è sottesa alla progettazione delle nuove tecnologie?
- Stiamo diventando sostituibili (replacement) o stiamo acquisendo nuove competenze? Cosa accade, per esempio, alla creatività umana?
- Quanto siamo disponibili a vivere in un mondo digitale ambiguo e perchè?
- Come sta cambiando la vita comunitaria e "tangibile" di fronte all'avvento della dimensione virtuale?



Grazie per l'attenzione!

gschiavo@fbk.eu

fabretti@fbk.eu

