

# Ma che cos'è questa intelligenza artificiale?

Angelo Montanari

Dipartimento di Scienze Matematiche, Informatiche e Fisiche

Università degli Studi di Udine

Pordenone, 26 maggio, 2023

# L'Intelligenza Artificiale (IA) all'Università di Udine

Classifica AI 2000 creata dalla Tsinghua University e dall'Associazione cinese di IA. Nel 2023, 4 docenti dell'Università di Udine fra i 2000 studiosi più influenti al mondo nella loro area di ricerca in IA. Sono Fabio Buttussi, Luca Chittaro, Angelo Montanari e Giuseppe Serra del Dipartimento di Scienze Matematiche, Informatiche e Fisiche.



# L'Intelligenza Artificiale (IA) è nello spirito del tempo (zeitgeist)

Intervento di Toby Walsh (NICTA, University of New South Wales, Australia), in una delle principali conferenze internazionali di IA (KR) nel 2016: **Will AI end jobs, wars or humanity?** Cape Town, SA, April, 2016:

- The Chief Economist of the Bank of England some years ago predicted **AI will destroy 50% of jobs in the UK.**
- Thousands of AI researchers signed an Open Letter predicting that AI could transform warfare and lead to an arms race of **“killer robots”**.
- S. Hawking and others have predicted that **AI could end humanity itself.**

Un paio di mesi fa: “Immediately suspend training of systems more powerful than gpt-4 for at least 6 months”, Future of Life Institute (E. Musk, S. Wozniak, Y. Bengio e molti altri)

# Intelligenza Artificiale: entusiasti e apocalittici

E' possibile stendere un elenco di **applicazioni** presenti e future dell'IA che copre tutte le lettere dell'alfabeto, dalle auto senza pilota alla bionica, dalla domotica alla (nuova) robotica industriale.

- **Entusiasti** (la filosofia postumanistica/transumanistica e il superamento dell'umano): Marvin Minsky (la società della mente), Raymond Kurzweil (la singolarità tecnologica).
- **Apocalittici**: Stephen Hawking, ma non solo.
- Nuove forme di **luddismo**: lotta violenta contro l'introduzione di nuove machine artificiali (la curiosa denigrazione dell'algoritmo).

# Né idolatrare né demonizzare

L'uomo costruisce gli **idoli** e poi li venera (Esodo 32), ma gli idoli (come le macchine, per quanto sofisticate) sono prodotti dell'essere umano.

Il problema può essere dimenticare l'origine dei sistemi artificiali intelligenti e idolatrarli (o demonizzarli).

Ciò non vuol dire che il **rapporto con tali macchine** sia facile:

- da sempre le macchine fanno delle cose che l'uomo non è in grado di fare (pensiamo alla rivoluzione industriale e ai mezzi di trasporto);
- il problema è sostituire e, spesso, migliorare le prestazioni dell'essere umano in compiti ritenuti da sempre di sua competenza esclusiva (**intelligenza artificiale: ragionamento e memoria**).

# Un vocabolario antropomorfo

- L'uso di un **vocabolario antropomorfo** nella descrizione delle caratteristiche e del funzionamento dei sistemi informatici
  - è particolarmente evidente nel caso dei sistemi di intelligenza artificiale (**intelligenza**, conoscenza, apprendimento, ragionamento),
  - ma si è verificato in misura più o meno rilevante in molti altri ambiti dell'informatica (**memoria**, comunicazione, interrogazione).
- **Ragioni** dell'uso di un vocabolario antropomorfo
  - Uomo/animale come modello in cibernetica e successivamente in diversi settori dell'informatica (Intelligenza Artificiale, Robotica, Bionica, ..).

# Norbert Wiener



- Norbert Wiener: *Cybernetics or Control and Communication in the Animal and the Machine*, MIT Press, 1962).

# Il rapporto uomo/macchina

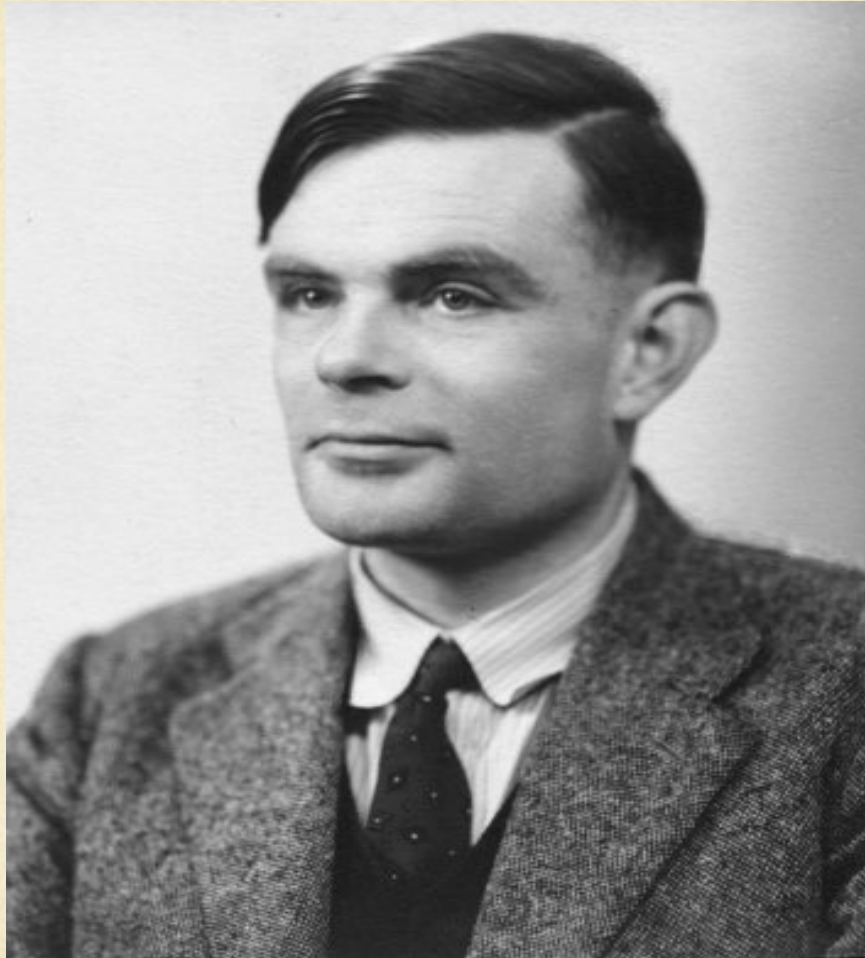
- Ogni **discorso** sulle proprietà "antropomorfe" delle macchine/calcolatori non riguarda tanto la macchina (il calcolatore) in sé, ma il modo in cui noi vediamo la macchina, e indirettamente noi stessi (ad esempio, Minsky rivendica la legittimità/utilità dell'uso di termini antropomorfici).
- Ciò vale, ad esempio, per la questione relativa alla **intelligenza/intenzionalità** delle macchine: parlare delle macchine è un modo (indiretto) per parlare di noi stessi (è una questione antropologica).



# Che cos'è l'Intelligenza Artificiale?

- Per quanto rilevante, l'Intelligenza Artificiale è da sempre un importante settore dell'informatica (era già presente negli scritti di Alan Turing, il padre riconosciuto dell'informatica).
- Lo specifico dell'informatica: la capacità di **modellare** e di **risolvere problemi**, dove la prima abilità è almeno tanto importante quanto la seconda.
- Il concetto chiave di **algoritmo**: una descrizione finita e non ambigua di una sequenza di operazioni che consente ad un certo agente di risolvere un determinato **problema**.

# Alan Turing



Computing Machinery and Intelligence, A. M. Turing, *Mind*, New Series, Vol. 59(236), October 1950, pp. 433-460.

# IA guidata dai modelli e IA guidata dai dati

In IA sono da sempre presenti **due filoni** principali: l'IA guidata dai modelli (IA simbolica) e IA guidata dai dati (IA sub-simbolica).

**IA simbolica:** rappresentazione della conoscenza e ragionamento automatico. Un caso paradigmatico: la pianificazione automatica.

**IA sub-simbolica:** mimare il comportamento del cervello umano e la sua complessa rete di neuroni interconnessi. I primi modelli di rete neurale risalgono agli anni '40 (il modello dei neuroni di McCulloch e Pitts, 1943).

Il caso dell'**elaborazione del linguaggio naturale:** dalla logica ai corpora (collezioni di grandi dimensioni di testi orali o scritti prodotti in contesti comunicativi reali).

# L'esplosione del machine/deep learning

Quali sono le ragioni dell'esplosione della ricerca e delle applicazioni del machine/deep learning?

Enorme **potenza di calcolo** ed enorme disponibilità di **dati**.

Il caso di **ChatGPT (Generative Pre-trained Transformer)**: un miliardo di parametri.

L'altra faccia della medaglia:

- costi economici;
- costi ambientali (consumo di energia / impatto ecologico).

# Trustworthy AI

**Human-in-the-loop:** un modello/sistema che prevede l'interazione col soggetto umano (ad esempio, nei processi di decisione).

Non sempre è possibile garantire tale condizione (applicazioni finanziarie e sistemi critici dal punto di vista della sicurezza).

In alcuni casi, è essenziale (diagnosi mediche «critiche»).

Obiettivo: **trustworthy AI**.

L'IA di cui ti puoi fidare, perché è in grado di rendere ragione /spiegare le conclusioni/previsioni formulate.

L'AI act europeo: come coniugare libertà e responsabilità nell'IA.

# A che punto siamo?

L'obiettivo trustworthy AI non è (ancora) stato raggiunto.

«ChatGPT è una promessa che deve essere accolta con estrema attenzione/prudenza» D. Pedreschi, Partenariato esteso FAIR (Future Artificial Intelligence Research).

ChatGPT non sa che cosa non sa, ossia non è in grado di dare una misura della robustezza/affidabilità delle sue risposte.

ChatGPT sta usando le persone, spesso a loro insaputa, per la validazione e l'affinamento delle sue capacità («fine tuning»)

# Come porsi nei confronti dell'IA?

La valutazione di un ricercatore di IA non può che essere **positiva**, pur nella consapevolezza di tutte le **criticità**.

Un **esempio** fra i tanti possibili (l'IA è ovunque, dalla guida autonoma ai chatbot, come ChatGPT): **early failure detection** e predictive maintenance (dagli impianti industriali e i sensori alla salute delle persone).

**I dati fotografano il passato**: il comportamento umano ripetitivo (dalle vacanze alla politica) è prevedibile/replicabile e, di conseguenza, influenzabile.

Elogio della **creatività**: dobbiamo esercitare il pensiero critico ed imparare ad esplorare nuove strade (futuro aperto)