

Intelligenza Artificiale: presente, passato e futuro

Angelo Montanari

Dipartimento di Scienze Matematiche, Informatiche e Fisiche

Università degli Studi di Udine

Malnisio Science Festival - Malnisio, 24 novembre, 2024

Indice degli argomenti

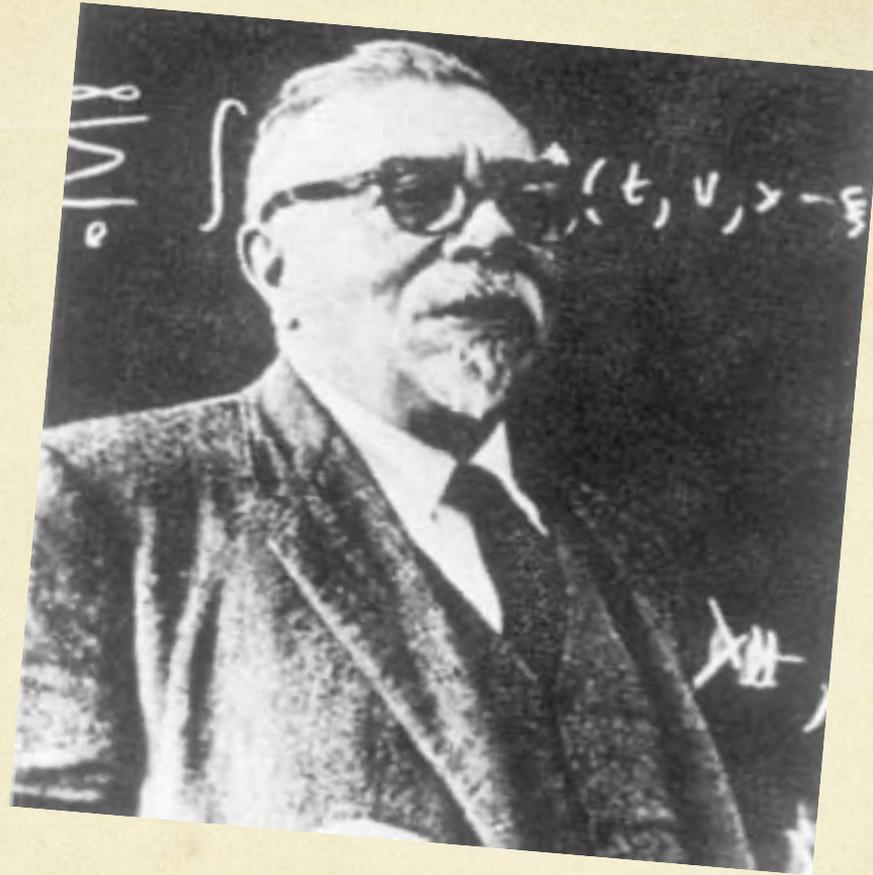
- Introduzione
- Ragionamento (la logica) e memoria (i dati)
- Intelligenza artificiale e esseri umani
- Considerazioni finali

Una questione preliminare

- Uso di un vocabolario antropomorfico nella descrizione delle caratteristiche e del funzionamento dei sistemi informatici
 - è particolarmente evidente nel caso dei sistemi di IA (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 tale vocabolario:
 - uomo/animale come modello di riferimento in cibernetica e, successivamente, in diversi settori dell'informatica (IA, Robotica, Bionica, ..).

Norbert Wiener



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

Che cos'è l'Intelligenza Artificiale?

- Per quanto rilevante, l'Intelligenza Artificiale (IA) è da sempre un settore dell'informatica (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** (un esempio di algoritmo molto familiare: una ricetta).

IA guidata dai modelli o 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 (di sistemi artificiali quali robot, satelliti, ..)

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 (Chat-GPT):** dalla logica ai corpora (collezioni di grandi dimensioni di testi orali o scritti prodotti in contesti comunicativi reali)

I sistemi di intelligenza artificiale

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).

La novità è che i sistemi di IA sono in grado di sostituire l'essere umano e, spesso, di migliorarne le prestazioni in compiti ritenuti da sempre di sua competenza esclusiva (intelligenza artificiale: ragionamento e memoria).

Quando i sistemi artificiali toccano la sfera dell'intelligenza siamo più sensibili e reagiamo in modo «intenso».

Il ragionamento automatico

Uso della logica (matematica) per la rappresentazione della conoscenza e il ragionamento automatico.

Teoria.

La dimostrazione automatica dei teoremi (theorem proving),
la verifica automatica delle dimostrazioni (proof checking).

Applicazioni.

- (i) la pianificazione automatica;
- (ii) la robotica collaborativa;
- (iii) il Business Process Management (BPM).

La pianificazione automatica

Pianificazione automatica: dati la situazione iniziale, un obiettivo e un insieme di vincoli, **costruire** un **piano** di azioni (se esiste) che consenta di raggiungere l'obiettivo rispettando i vincoli dati.

Sintetizzare un **controllore** in grado di eseguire tale piano.

Monitorare l'esecuzione del piano ed intervenire (adattando il piano - ripianificazione) qualora si rilevino situazioni impreviste/indesiderate.

Si usa per l'automazione dei processi di produzione, ma non solo.

La robotica collaborativa

La robotica collaborativa combina le potenzialità e le capacità delle persone e dei robot per portare a termine, in maniera semi-automatica, dei compiti che risulta impossibile realizzare in maniera completamente automatica.

Che cosa sono i robot collaborativi (**cobot**)?

Sono dei robot industriali (ma vi sono applicazioni anche nell'ambito dell'assistenza domiciliare e della domotica) di nuova generazione progettati per lavorare assieme all'essere umano, fianco a fianco e in sicurezza, senza la necessità di barriere o gabbie protettive che separino gli uni dagli altri.

Il Business Process Management (BPM)

Mette a disposizione una molteplicità di strumenti per la **gestione dei processi aziendali**.

Strumenti che consentano di scoprire, rappresentare ed analizzare i compiti sottostanti ai processi aziendali (**process mining**).

Strumenti che consentano di **visualizzare** i processi aziendali (diagrammi).

Strumenti per l'automazione dei processi / della gestione dei processi aziendali (**workflow engine**).

Strumenti per la **simulazione** e il **testing** di processi aziendali.

La memoria

La memoria è una componente essenziale dell'intelligenza (umana).

La dimensione temporale dell'intelligenza: memoria e **temporalità**.

Il passato: i dati.

Il futuro: le previsioni e i piani.

L'apprendimento (automatico) sulla base dei dati: non può esserci apprendimento senza dati.

Intelligenza come capacità di **classificazione** dei dati.

La memoria e la rete

Dalla memoria del calcolatore alla rete.

La **rete**: un'enorme memoria distribuita e condivisa.

Una **memoria senza soggetto** (una memoria impersonale).

Quali sono le ragioni dell'esplosione della ricerca e delle applicazioni del machine/deep learning (IA guidata dai dati)?

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

Le applicazioni dell'IA guidata dai dati

Manutenzione predittiva (per prevenire il fermo macchina): monitoraggio in tempo reale del comportamento di un sistema per il rilevamento anticipato di potenziali guasti o malfunzionamenti

.. dall'ambito della produzione industriale a quello della salute umana

Tecniche di profilazione di utenti/clienti/elettori..

Upselling predittivo

Le molteplici applicazioni dell'IA generativa

Intelligenza ed essere umano

L'intelligenza (umana) consiste unicamente di ragionamento e memoria?

Intelligenza e «umanità»: la coscienza è un tratto distintivo dell'intelligenza o dell'umanità (o dell'intelligenza dell'essere umano)?

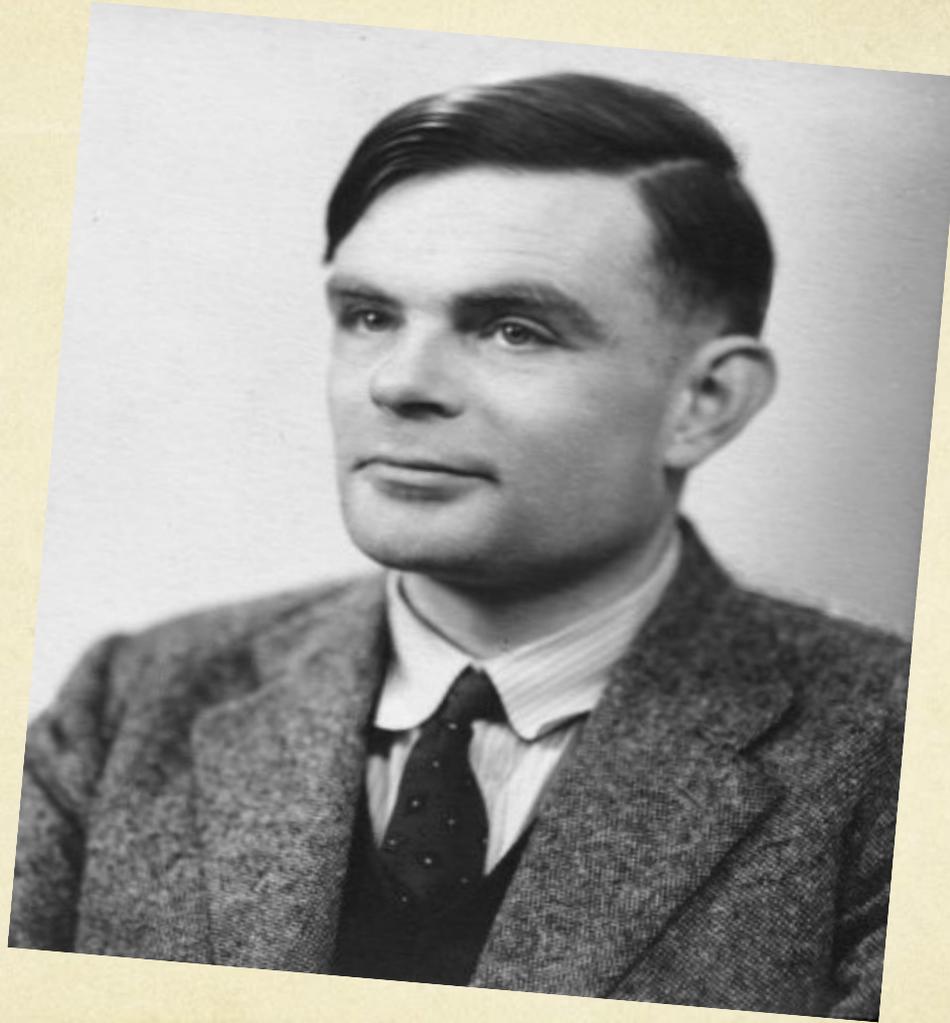
L'intelligenza umana (naturale) è un'intelligenza soggettiva («in prima persona») ossia l'intelligenza di un soggetto dotato di (auto)coscienza e di intenzioni.

Il test di Turing

Il **test di Turing** (o gioco dell'imitazione): una macchina può essere definita intelligente se riesce a convincere una persona che il suo comportamento, dal punto di vista intellettuale, non è diverso da quello di un essere umano medio



Alan Turing



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

Intelligenza e linguaggio

Il Test di Turing assume un legame molto stretto tra intelligenza e **linguaggio** (naturale): l'intelligenza si manifesta nell'interazione/comunicazione attraverso il linguaggio.

Il linguaggio occupa un ruolo fondamentale nel rudimentale **sistema esperto ELIZA** (psicanalista digitale) proposto da Weizenbaum negli anni '60.

La comprensione di testi in linguaggio naturale è al centro del famoso **esperimento mentale della stanza cinese** di Searle (impossibilità per una macchina di manifestare l'**intenzionalità** che caratterizza gli esseri umani e, sia pure in forme diverse, gli animali).

All'ambito dell'elaborazione del linguaggio naturale appartiene anche l'universo dei sistemi conversazionali (Chatbot), dei quali **ChatGPT** è uno dei rappresentanti più noti.

Intelligenza e corporeità

Il modello di intelligenza sotteso al Test di Turing è un modello astratto/disincarnato dell'intelligenza.

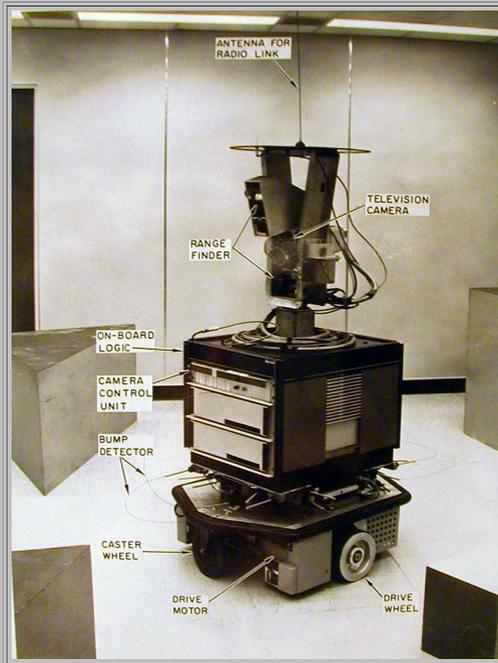
Una delle acquisizioni più importanti della ricerca in IA degli ultimi decenni è la consapevolezza del ruolo cruciale che gli organi di senso svolgono nell'interazione dell'uomo col mondo e della conseguente impossibilità di un'intelligenza (artificiale) priva di "corporeità". Ciò ha portato allo sviluppo di un rapporto sempre più stretto tra IA ("cervello senza corpo") e robotica ("corpo senza cervello").

Per paradossale che possa suonare, per avvicinarsi all'intelligenza umana l'IA deve diventare un'intelligenza incarnata.

Shakey ('60/'70), TMSUK 04 ('80/'90) e Justin ('09-)



Shakey the Robot - SRI (1968)



Autonomia delle macchine

Nei sistemi robotici e, più in generale, nei sistemi di IA, non sempre è possibile (spesso non lo è) garantire la condizione «**human in the loop**» (un modello/sistema che preveda l'interazione col soggetto umano, ad esempio, nei processi di decisione).

Non confondere l'incapacità di prevedere/controllare il comportamento di una macchina con un'attività creativa della macchina.

Più una macchina diventa complessa/potente più risulta difficile controllarne il comportamento.

Il problema dell'**autonomia** e della responsabilità delle macchine. Si pensi alle auto a guida autonoma, ma anche all'uso dell'IA in ambito militare.

Intenzionalità delle macchine

Coscienza artificiale: una questione antropologica.

Per poter ragionare di coscienza delle macchine (coscienza artificiale) occorre preliminarmente stabilire **cos'è la coscienza umana** (naturale).

Posizioni fortemente divergenti.

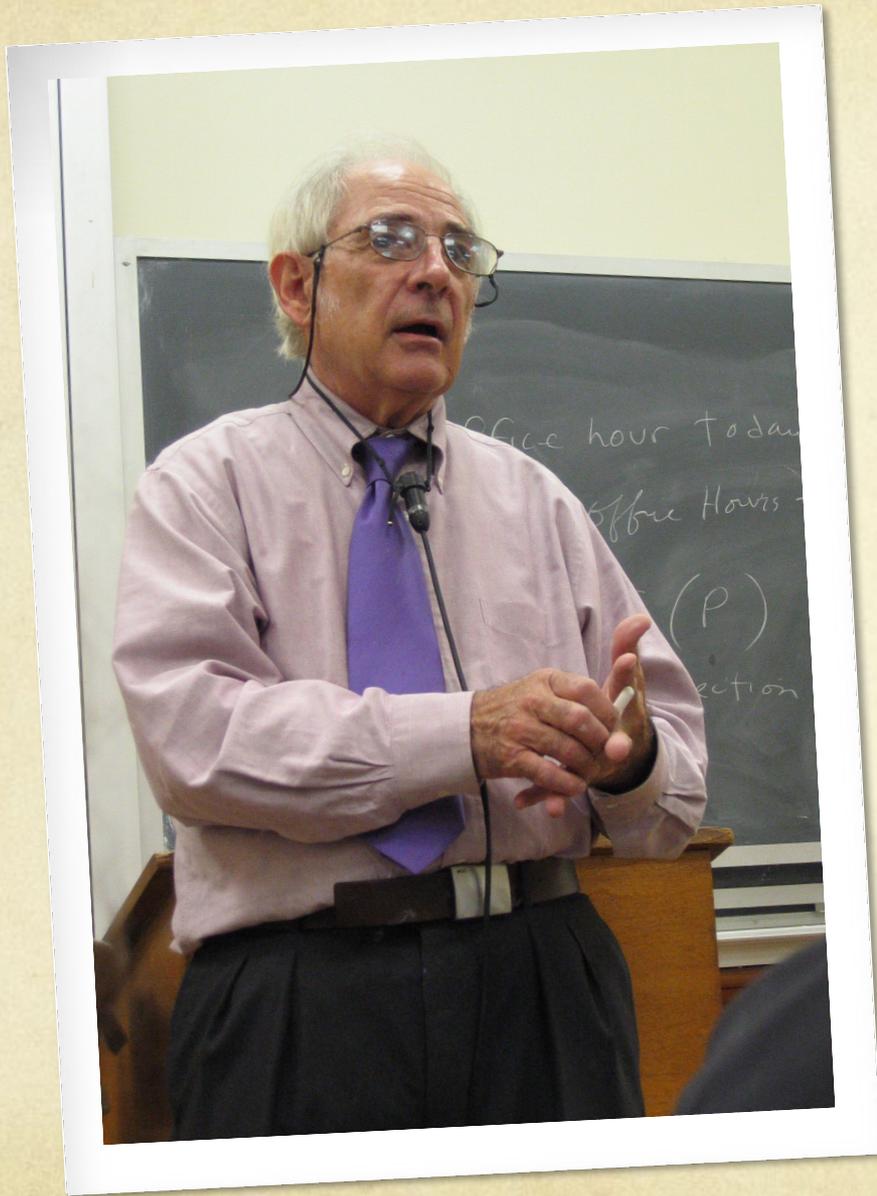
Marvin Minsky: la coscienza ha più a che fare con le cose semplici/banali che non con le cose complicate/profonde.

John Searle: l'intenzionalità come caratteristica distintiva della coscienza e l'impossibilità per un sistema di IA, come quelli sin qui sviluppati, di possederla.



Marvin Minsky

The Society of Mind,
TOUCHSTONE BOOK 1988
(tr. it. La società della mente,
Biblioteca Scientifica 10,
Adelphi, 1989).



John Searle

John. Searle, *Minds, brains, and programs*, *Behavioral and Brain Sciences*, vol. 3, 1980.

La stanza cinese di Searle

Tratto distintivo della coscienza: l'intenzionalità.

Impossibilità per una macchina di manifestare l'intenzionalità che caratterizza gli esseri umani e, sia pure in forme diverse, gli animali.

Tesi: L'esecuzione di un programma (algoritmo) su un dato input non è mai di per se stessa una condizione sufficiente per l'intenzionalità.

La **dimostrazione** (un esperimento mentale): Searle immagina di sostituire un agente umano al calcolatore nel ruolo di esecutore di una specifica istanza di un programma e mostra come tale esecuzione possa avvenire senza forme significative di intenzionalità.

Alcune considerazioni

L'affermazione dell'irriducibilità dell'intenzionalità all'esecuzione di programmi su particolari input ha alcune importanti conseguenze:

- impossibilità di spiegare le modalità con le quali il cervello produce l'intenzionalità attraverso l'esecuzione di programmi su particolari input
- ogni meccanismo in grado di produrre intenzionalità deve avere abilità di tipo causale pari a quelle del cervello

Problemi (irrisolti): cosa differenzia il caso in cui la persona comprende il testo (inglese) da quello in cui non vi è alcuna comprensione (cinese)? Questo qualcosa può (se sì, come) essere trasferito ad un macchina?

Intenzionalità derivata

Il legame tra **intenzionalità** (umana) e capacità di creare artefatti: la creazione di artefatti è per l'uomo un modo per estendere la propria intenzionalità

“i nostri strumenti – osserva Searle – sono estensioni dei nostri scopi/intenzioni”

L'intenzionalità trova modo di esprimersi nella creazione di artefatti, dei quali costituisce, in un certo senso, la causa finale (**intenzionalità derivata**), e, successivamente e in modo diverso, nel loro utilizzo, ma non si trasferisce ad essi.

Come parlare di intenzionalità derivata in sistemi artificiali come Chat GPT? Focus dal ragionamento alla memoria.

Il problema dell'affidabilità

In molti casi, quali le applicazioni finanziarie e diversi sistemi critici dal punto di vista della sicurezza (velivoli, centrali per la produzione di energia, ..), la condizione «human in the loop» non può essere soddisfatta

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

Una possibile strada: **integrazione** di IA simbolica e IA sub-simbolica.