Oracoli o medici: le sfide epistemologiche dell'AI in sanità

Ciro De Florio

Facoltà di Economia, Humane Technology Lab
Università Cattolica di Milano



AIEOP
ASSOCIAZIONE ITALIANA EMATOLOGIA ONCOLOGIA PEDIATRICA

AIEOP MEDICI
LECCE, 12-14 ottobre 2026
HILTON GARDEN INN

Disclosures of Name Surname

Company name	Research support	Employee	Consultant	Stockholder	Speakers bureau	Advisory board	Other

Agenda: di che cosa parleremo oggi

- Che cosa è — e come funziona — un sistema di AI:
  qualche parola chiave
- Chiediamo al dottore:
  fiducia epistemica ed epistemologia della testimonianza
- Prevedere e inferire, capire e spiegare
- Per concludere: *manifesto* per un'alleanza epistemica

Di cosa NON parleremo

- Impiego di sistemi di AI nell'organizzazione e nel management sanitario
- Doctor AI: educazione all'AI da parte di pazienti e famiglie
- Etica* dell'AI

Che cosa è un sistema di AI

«L'AI mira alla costruzione di macchine in grado di svolgere compiti che, **se fossero svolti da esseri umani**, richiederebbero intelligenza» (McCarthy)

I approccio

Programmare una macchina costruendo algoritmi in cui **impacchettare** la nostra conoscenza

II approccio

Costruire algoritmi in grado di *apprendere* a partire da dati in maniera autonoma

Che cosa è un sistema di AI

Machine Learning:

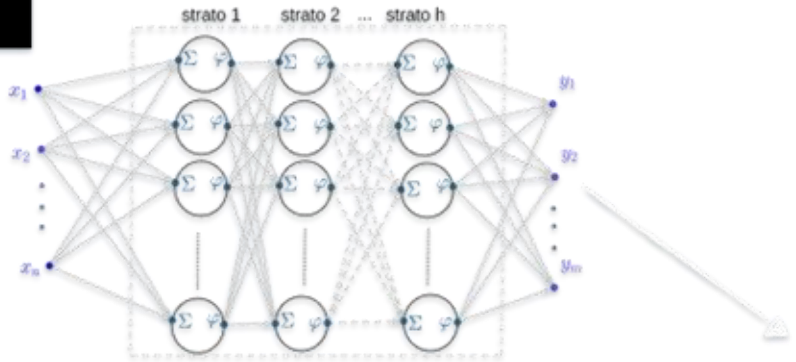
«Lo scopo dei modelli di ML è quella di *imparare* in maniera *autonoma* una funzione che *predica* proprietà *sconosciute* da pattern osservati in precedenza» (Hardt and Recht 2022, 16)

- apprendimento;
- autonomia;
- addestramento;

Che cosa è un sistema di AI

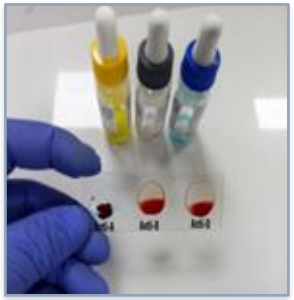
Un esempio:





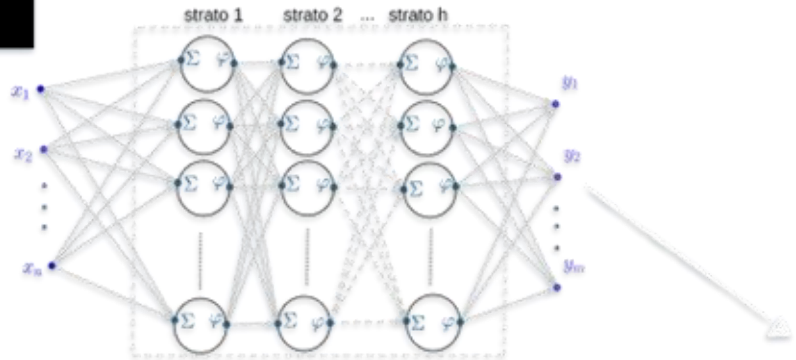
Che cosa è un sistema di AI

Un esempio più articolato:









Fiducia epistemica ed epistemologia della testimonianza

Perché ci fidiamo dei medici?

- Autorità epistemica
  - Ragioni **intrinseche**
    - **Teorie** sul funzionamento del corpo umano
    - **Comprensione** dei fenomeni
    - …
  - Ragioni **estrinseche**
    - Comunità scientifica
    - Autorevolezza personale
    - …

Ci sono ragioni per fidarsi / non fidarsi dei sistemi di ML in ambito medico?

Prevedere e inferire, capire e spiegare

- Fiducia epistemica basata sull'**accuratezza del risultato** e **comprensione dei meccanismi** del dominio in oggetto
- Per comprendere relazioni **causali** complesse abbiamo bisogno di **teorie**
- Il modello addestrato sui dati — non importa quanto accurato — è **a-teorico**
- Quindi, affidabilità senza comprensione: concezione *oracolare*

Prevedere e inferire, capire e spiegare

- Il nostro *prompt* sia *A* (l'esito di un determinato esame, per esempio);
- La macchina restituisce *B* (un'ipotesi diagnostica)

$$ML: A \vdash B$$

Prevedere e inferire, capire e spiegare

- Il nostro *prompt* sia *A* (l'esito di un determinato esame, per esempio);
- La macchina restituisce *B* (un'ipotesi diagnostica)

$$ML: A \vdash B$$

- Ma *A* in realtà appartiene a un dominio più generale di informazioni, di tipo $\mathfrak{A}$ e *B* appartiene a un dominio più generale di diagnosi, di tipo $\mathfrak{B}$
- Per esempio, *A* può essere un determinato esame del sangue e *B* una determinata diagnosi che riguarda la presenza di un certo valore ematico

Prevedere e inferire, capire e spiegare

- Il nostro *prompt* sia *A* (l'esito di un determinato esame, per esempio);
- La macchina restituisce *B* (un'ipotesi diagnostica)

$$ML: A \vdash B$$

- *B* è tanto più **plausibile** quanto esiste un **apparato teorico** che connette domini di tipo $\mathfrak{A}$ con informazioni di tipo $\mathfrak{B}$
- Se la connessione teorica tra $\mathfrak{A}$ e $\mathfrak{B}$ è debole allora l'accettazione di *B* diventa problematica

Prevedere e inferire, capire e spiegare

- Il nostro *prompt* sia *A* (l'esito di un determinato esame, per esempio);
- La macchina restituisce *B* (un'ipotesi diagnostica)

$$ML: A \vdash B$$

Quindi, la teoria sul funzionamento del sistema, la **scienza medica**, serve **non perché spiega direttamente *B*** ma perché rende plausibili risultati simili a *B* dati input simili ad *A*

manifesto per un'alleanza epistemica

Machine side

- Software di ML dedicati, con *theory injection*
- Cura e attenzione al set di addestramento: dati DOCG
- …

Human side

- Conoscere architetturalmente il funzionamento dei modelli
- Continuare a sviluppare **teoria**
- Negoziare il trade-off tra plausibilità e informazione

Grazie per la vostra attenzione

ciro.deflorio@unicatt.it