



## Scuola Superiore di Catania

*Corso specialistico*

a.a. 2024-2025

### **Machine Learning and Knowledge Building. Storia, epistemologia ed etica nell'intelligenza artificiale**

*Machine Learning and Knowledge Building. History, epistemology and ethics in artificial intelligence*

#### Obiettivi formativi

##### Competenze:

Offrire agli studenti e alle studentesse competenze critiche in materia di storia ed epistemologia dell'intelligenza artificiale con particolare riguardo alle tecniche basate sull'apprendimento. Mettere studenti e studentesse in condizione di definire le corrette scelte politiche ed etiche necessarie per mantenere intatti i nostri standard democratici nella tutela dei diritti di cittadini e comunità, con particolare riguardo alla presa di decisione automatica in contesti incerti e contingenti relativi alle situazioni concrete di vita delle persone

##### Contenuti:

Uno sguardo storico: • Il problema di Turing • IA e cibernetica • Servomeccanismi e calcolatore digitale • IA e processo creativo La datificazione digitale come nuovo sistema di esternalizzazione della memoria e le sue conseguenze epistemiche e politiche;

La distinzione tra intelligenza artificiale basata sul reasoning e quella basata sul learning;

Le aree in cui si usa l'intelligenza artificiale che sono sensibili dal punto di vista della tutela dei diritti collettivi e individuali: • La presa di decisione automatica o il supporto alla decisione nei contesti incerti • Il riconoscimento di immagini, volti, emozioni • La produzione di contenuti multimediali in conseguenza di un prompt governato dagli esseri umani Le tecniche inferenziali adoperate per estrarre regolarità e previsioni dai dati: • Inferenza induttiva • Clusterizzazione • Riconoscimento • Produzione statistica di contenuti;

Le questioni politiche irrisolte in conseguenza delle tecniche usate per anticipare comportamenti e predire situazioni: è possibile giustificare le inferenze sul futuro senza collegarvi una spiegazione? Siamo sicuri di trovarci in un contesto privo di discrezionalità nell'attività di interpretazione algoritmica dei dati? Come è possibile costruire delle funzioni obiettivo controllabili e condivisibili che



rispondano alla necessità di mantenere intatti i livelli di tutela collettivi previsti dalle regole democratiche? Come far valere il principio della fairness algoritmica senza rinunciare a esplicitare le regole applicate dai sistemi tecnici?

Individuare gli ambiti della delega algoritmica che non creino difficoltà epistemiche e politiche e dove la previsione non abbia in sé il carattere di prescrizione: • Apprendimento automatico e decisione umana • Medicina e IA • Governance algoritmica • Armi autonome e controllo umano significativo • La moralizzazione dei sistemi di IA • Etica dei veicoli autonomi.

### Metodologia didattica

- Didattica frontale
- Seminari partecipativi

### Modalità della verifica finale di apprendimento

Esame finale e stesura di un paper e/o presentazione.