

PATHS

Summer School

10-12 giugno 2024



PATHS

Summer School

10-12 giugno 2024

Etica dell'intelligenza artificiale

Adriano Fabris



Una definizione di AI

Fra le tantissime che ci sono...

Artificial intelligence (AI) refers to systems that display intelligent behavior by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals.

AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications).

A Definition of AI: Main Capabilities and Disciplines: cfr.

<https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top>

Che cosa c'entra l'etica con l'AI? – 1 –

- L'etica è la riflessione sui criteri e i principi che sovrintendono ai nostri comportamenti
- Di questi criteri e principi abbiamo bisogno per fare delle scelte
- Essi possono essere criteri e principi individuali (riferiti specificamente a un individuo, anche in polemica con la società) oppure condivisi (da un gruppo, da una società). I due sensi della parola greca *ethos*
- La filosofia fin dalle sue origini cerca di stabilire e giustificare criteri e principi condivisibili da tutti
- Lo fa perché vuole porsi come antidoto all'ingiustizia e alla violenza. Il trauma della morte di Socrate

Che cosa c'entra l'etica con l'AI? – 2 –

- I criteri e i principi individuati dall'etica sono i criteri e i principi dell'agire
- Ora – questa è la novità con cui dobbiamo fare i conti – l'agire non è più soltanto l'agire umano: ora capaci di azione – soggetti di azione – sono anche programmi ed entità artificiali. Si parla perciò di «agenti artificiali morali»
- Se ne parla perché questi agenti sono dotati di un certo grado di autonomia (si veda la definizione di AI nella slide 3)

Tre tipologie di etica relativa all'AI

Bisogna dunque porsi domande etiche anche riguardo agli agenti artificiali. Ciò porta all'identificazione di 3 tipologie di etica dell'AI:

1. Etica inserita nei software AI
2. Etica nelle modalità di interazione tra esseri umani e agenti artificiali
3. Etica nelle interazioni fra esseri umani, entità artificiali e ambiente in generale

Tre questioni etiche da affrontare concretamente

Le 3 tipologie di etica relativa all'AI richiedono di essere affrontate dall'essere umano in questi 3 modi:

1. Addestramento etico delle macchine
2. Educazione etica degli esseri umani
3. Condivisione di valori universali per la regolamentazione delle relazioni

Tre linee di approfondimento della ricerca filosofica

Tutto ciò porta a definire 3 linee di ricerca nell'ambito dell'etica relativa all'AI:

1. Etica *nella* AI
2. Etica *della* AI
3. Analisi, giustificazione e promozione dei valori universali in gioco

Approcci didattici e comunicativi

- È questo il sottotitolo del mio intervento
- All'interno di una riflessione più ampia sull'etica relativa all'AI si delineano questioni di fondo che impattano, fra le altre, sugli ambiti della didattica e della comunicazione
- Vale la pena di considerare schematicamente alcune di tali questioni

Questioni di fondo che concernono la didattica – 1 – L’impatto dell’AI sulla didattica

- L’impatto nell’uso da parte degli studenti di certi dispositivi
- L’esempio di ChatGPT e del suo utilizzo anche improprio
- L’impatto di certi programmi sulle forme d’insegnamento
- Il modello di domanda e risposta nel caso dei LLM: si ha a che fare davvero con qualcosa (o qualcuno) di simile all’essere umano?
- In tutto ciò qual è il ruolo dell’insegnante?

Questioni di fondo che concernono la didattica – 2 – L'impatto dell'AI sulla mentalità comune

- Il tema di fondo è: si tratta di un affiancamento dell'essere umano o della sua sostituzione?
- Da parte dello studente: tendenza alla sostituzione (di ciò che fa e di ciò che è), per motivi di efficienza e di comodità. La conseguenza è quella del deskilling: impoverimento delle competenze
- Da parte del docente: tentazione della sostituzione, anche in questo caso, subordinandosi a un altro modello di sapere (ritenuto migliore). La conseguenza, fra l'altro, è la (definitiva) proceduralizzazione dell'insegnamento

Antidoti

- Bisogna evitare, fra l'altro, sostituzione, deskilling, servitù volontaria (secondo l'espressione di De la Boetie)
- Bisogna evitare però, anche, un atteggiamento unicamente proibizionista
- L'insegnante fa quello che la macchina non può fare: può insegnare non solo come funziona la macchina, ma qual è il senso e quali sono le conseguenze del suo funzionamento
- Questo è il compito dell'etica dell'AI applicata alla didattica

Questioni di fondo che concernono i modi di comunicare questa nuova situazione

- Come comunicare tutto questo? Alcune possibili risposte:
- Comunicare correttamente l'AI e le sue potenzialità significa collocarsi giustamente in rapporto a essa
- Significa evitare l'effetto-Disney: sia l'antropomorfizzazione delle macchine che l'artificializzazione degli esseri umani
- Significa sostenere la prospettiva di una human-centered AI. È la prospettiva europea; è la prospettiva del governo italiano

Human-centered AI

- Questa prospettiva è esposta in:
- Documenti europei (soprattutto in quello sulla Trustworthy AI e nell'AI Act)
- Varie iniziative del Governo italiano
- Essa poi fa riferimento a valori condivisi, a criteri etici: soprattutto quelli della trasparenza, dell'attenzione per l'autonomia dell'essere umano, del rispetto della sua privacy, dell'esplicabilità dei processi, dell'equità di accesso ai programmi, della responsabilità e della sicurezza nel loro uso

Questo, appunto, è ciò che va insegnato e comunicato: realizzando così, concretamente, un'etica dell'AI

Grazie per l'attenzione!