Attenti all'AI: aspetti etici

leonida iannantuoni

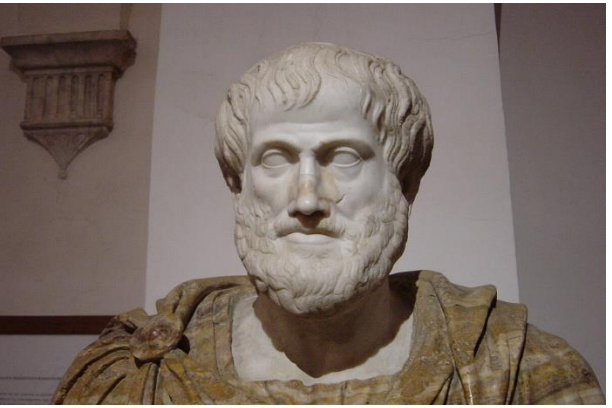
ETICA

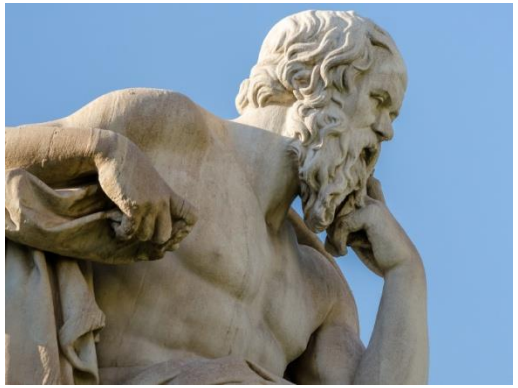
L'etica è una branca della filosofia che studia i principi e i valori che guidano le azioni umane, distinguendo tra ciò che è giusto e ciò che è sbagliato.

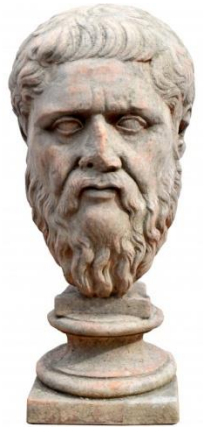
L'etica è, quindi, sia un insieme di norme e di valori che regolano il comportamento dell'uomo in relazione agli altri, sia un criterio che permette all'uomo di giudicare i comportamenti, propri e altrui, rispetto al bene e al male.

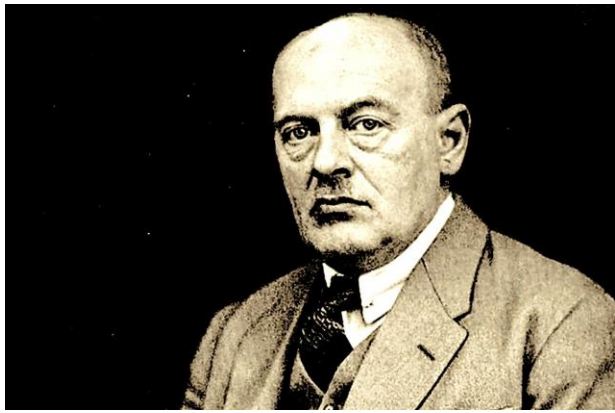
ETICA

Hanno contribuito allo sviluppo dell'etica: Socrate, Platone, Aristotele, Kant e gli utilitaristi come Bentham e Mill, più recentemente Max Scheler, Beauchamp e Childress.





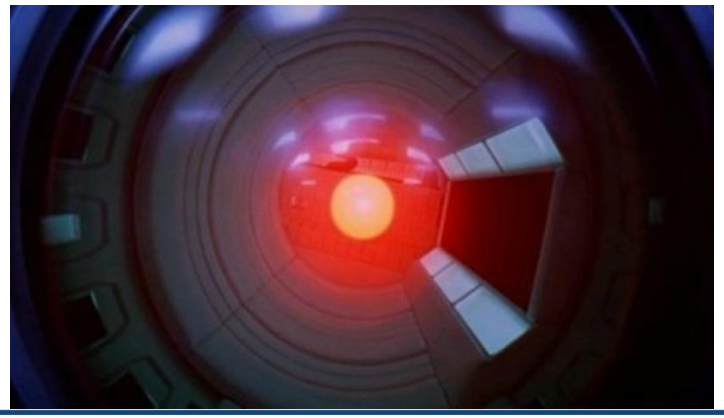




Etica dell'AI

L'AI funziona in base a come viene progettata, sviluppata, addestrata, messa a punto e utilizzata.

WARGAMES	2001 ODISSEA NELLO SPAZIO	TERMINATOR



Etica dell'AI?

Esempi di questioni etiche

responsabilità e privacy
equità
spiegabilità
solidità
trasparenza
sostenibilità ambientale

inclusione
agire morale
allineamento dei valori
responsabilità
uso improprio

Principi di etica dell'AI

la comunità accademica fa leva sul Belmont Report come mezzo per guidare l'etica nell'ambito della ricerca sperimentale e dello sviluppo algoritmico.

Dal Belmont Report emergono 3 principi fondamentali:

Rispetto delle persone

Beneficenza

Giustizia

Principi di etica dell'AI

Rispetto delle persone:
Questo principio sancisce l'autonomia delle persone.
Obbliga i ricercatori a proteggere le persone fragili a causa di una serie di circostanze come malattie, disabilità mentali, età, ecc.

Principi di etica dell'AI

Beneficenza:

Questo principio, deriva dal giuramento ippocratico "primum non nocere".

Applicato all'AI deve far sì che gli algoritmi, al di là delle intenzioni, non possano amplificare i pregiudizi su razza, genere, tendenze politiche, ecc.

Principi di etica dell'AI

Giustizia:

Questo principio affronta questioni come l'equità e l'uguaglianza.

Il Belmont Report offre cinque modi per distribuire oneri e vantaggi, ovvero:

- **Equa ripartizione**
- **Esigenza individuale**
- **Impegno individuale**
- **Contributo sociale**
- **Merito**

Bias e discriminazioni

Bias e discriminazioni presenti in una serie di applicazioni, software di riconoscimento facciale e/o algoritmi dei social media, hanno sollevato dubbi etici sull'uso dell'AI portando le aziende a prendere coscienza dei rischi.

L'IBM (2020) ha abbandonato i prodotti di riconoscimento facciale e analisi per uso generale, specificando che: "*IBM si oppone fermamente all'uso di qualsiasi tecnologia, compresa quella di riconoscimento facciale offerta da altri fornitori, per la sorveglianza di massa, la profilazione razziale, le violazioni dei diritti umani e delle libertà fondamentali, o qualsiasi scopo che non sia coerente con i nostri valori e i nostri principi di fiducia e trasparenza*".

Principi chiave

I principi chiave per l'uso etico dell'AI includono:
trasparenza,
equità,
responsabilità,
privacy,
supervisione umana.

Ulteriori principi: sicurezza, robustezza tecnica, non discriminazione, sostenibilità.
Tali principi guidano sviluppo ed applicazione dell'IA a garanzia che sia a beneficio dell'umanità e minimizzi i danni.

Principi chiave

Trasparenza e intelligibilità:

I processi decisionali dell'IA devono essere comprensibili agli esseri umani per identificare errori e bias.

Equità e non discriminazione:

L'IA deve trattare tutti gli individui in modo imparziale, ed evitare discriminazioni basate su etnia, genere, età, ecc..,

Responsabilità (Accountability):

Sviluppatori ed utenti sono responsabili di azioni e decisioni prese dai sistemi di IA, con meccanismi di supervisione e rendicontazione.

Principi chiave

Privacy e sicurezza dei dati:

I sistemi di IA devono proteggere i dati personali, richiedere il consenso informato e rispettare le normative sulla privacy.

Supervisione umana:

Deve esserci un intervento umano nei cicli decisionali, specie in contesti critici, per garantire che le decisioni finali siano prese da una persona e non totalmente automizzate

Robustezza e sicurezza:

I sistemi di IA devono essere sicuri, affidabili e resilienti per ridurre al minimo i danni involontari e prevenire attacchi informatici.

Principi chiave

Benessere sociale e ambientale:

Lo sviluppo dell'IA deve contribuire al benessere sociale, alla prosperità e alla sostenibilità ambientale.

Promozione dell'autonomia umana:

L'IA non deve compromettere l'autonomia e la libertà degli esseri umani di stabilire i propri standard e le proprie norme.

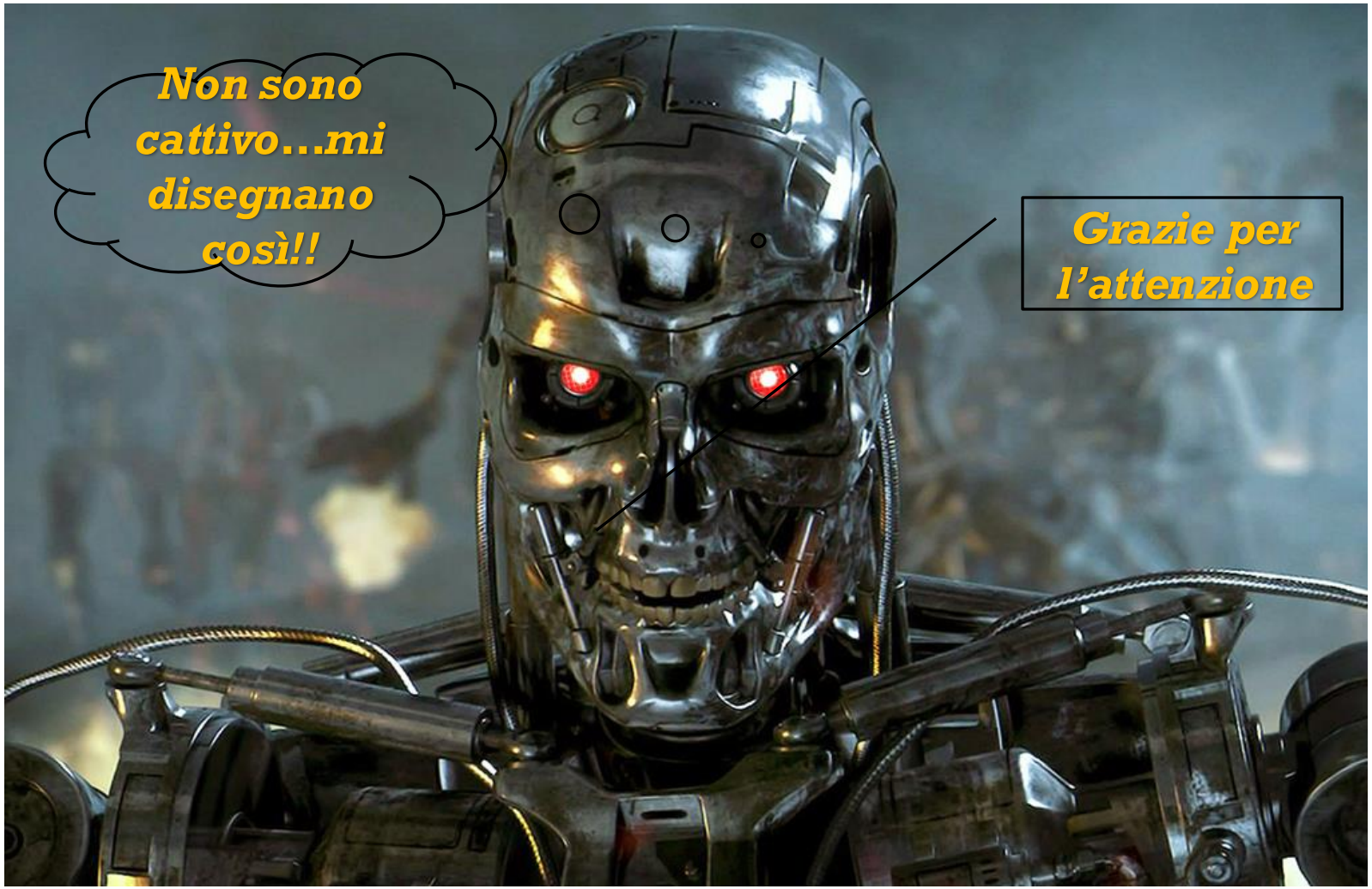
Organizzazioni che promuovono l'etica dell'AI

AlgorithmWatch: org. no profit, si concentra su algoritmi e processi decisionali spiegabili e tracciabili.

AI Now Institute: org. no profit della New York University studia le implicazioni sociali dell'AI.

DARPA: Defense Advanced Research Projects Agency del Dip. della Difesa degli USA si concentra sulla promozione dell'AI spiegabile e della ricerca.

CHAI: il Center for Human-Compatible Artificial Intelligence è una cooperazione di vari istituti e università che promuove un'AI affidabile.


Non sono cattivo…mi disegnano così!!
Grazie per l'attenzione