

Etica e intelligenza artificiale

Alessio Vaccari

IA E VITA QUOTIDIANA: LA NUOVA NORMALITÀ È GIÀ QUI

- L'intelligenza artificiale sta **trasformando radicalmente** la nostra società, le nostre attività professionali e la nostra vita quotidiana.
- E' **disseminata** in molti oggetti d'uso comune: **bancomat; 2) smartphone, 3) algoritmi che governano Google; Netflix; filtri antispam; 4) assistenti vocali: Siri, Alexa; 5) agenti artificiali come** le autovetture senza conducente alla robotica sociale (Sophia; Paro);
 - sito Robots dell'Institute of Electrical and Electronics Engineers:
<https://robotsguide.com>

IA E VITA QUOTIDIANA: LA NUOVA NORMALITÀ È GIÀ QUI

- Le nuove tecnologie digitali consentono pratiche **estremamente positive**:
 - 1) **citizen science/scienza partecipata**: le competenze e l'intuizione umana vengono sfruttate su larga scala (WIKIPEDIA; FOLDIT);
 - 2) **democrazia partecipativa online**: AVAAZ (aiuta le persone a mobilitarsi su temi globali, Me Too; Occupy Wall Street, Fridays for Future); CHANGE.ORG: piattaforma di petizioni on-line.
- Le nuove tecnologie **suscitano preoccupazione**:
 - 1) **mappano** la nostra vita (la misurano e la controllano), trasformano la nostra capacità di attenzione e di relazionarci agli altri;
 - 2) non solo **rispondono ai desideri** delle persone (conoscenza, intrattenimento, connessione), **ma plasmano anche lo spazio della scelta**;
 - 3) collaborano con il capitalismo – sono sviluppate dietro la pressione di fare maggiori profitti a partire dalle nuove risorse costituite dai **dati personali del corpo e della mente (dati biometrici; dati sui nostri consumi)**; il lavoro che scompare;
- Cambiamento **qualitativo** e accelerato. **Ruolo della filosofia**: (a) comprensione; (b) guida per il cambiamento: quale può essere un futuro buono per gli esseri umani?

CHE COS'È L'INTELLIGENZA ARTIFICIALE

1. Un programma scientifico che studia la possibilità di costruire **macchine intelligenti** ("Computing Machinery and Intelligence» 1950 di Alan Turing). Un progetto teorico che esplora la **natura della mente** e i processi cognitivi.
2. L'**evoluzione tecnologica** del computer, oggi presente in una vasta gamma di dispositivi, dagli smartphone agli elettrodomestici smart.
3. L'ecosistema di **Internet**: componenti tecniche; attori umani; regole; interazioni dinamiche.
4. Lo sviluppo di tecniche come il **machine learning**: un ramo dell'IA in cui le macchine imparano dai dati senza essere programmate con regole fisse per svolgere un singolo compito. Ai generativa: genera nuovi contenuti (ChatGPT).
 - Non imita l'**intelligenza umana** - cioè non pensa in modo simbolico – ma azzecca le risposte a partire da regolarità statistiche. E' **intelligente nel risultato e non nella produzione**. E' resa possibile dalla enorme quantità di dati con cui l'abbiamo nutrita.

CHE COS'E' L'ETICA

Etica quotidiana: una sfera del nostro linguaggio, delle nostre emozioni e dei nostri concetti (dovere, virtù, giusto, autonomia, libertà, ecc.) che utilizziamo per descrivere il mondo e i comportamenti umani.

Etica filosofica e teorie morali: utilitarismo (Bentham; Mill), deontologismo (Kant), etica della virtù (Aristotele; Hume).

Etica filosofica e progetto democratico umanistico: promuovere lo sviluppo autonomo e creativo del sé, accessibile al maggior numero possibile di persone. (Hume; Mill; Dewey).

IA: TRE QUESTIONI MORALI

L'Intelligenza Artificiale può davvero prendere decisioni al posto nostro? Una riflessione su efficienza, affidabilità e fiducia

I sistemi artificiali rendono più **efficiente** il lavoro dei ricercatori, medici, ecc. (maggiore velocità per quesiti che implicano grandi volumi di dati e minore costo), ma **sono affidabili** quanto quelle di un agente esperto? Il caso della medicina.

Architetture della scelta. Consigli su Misura o Controllo Invisibile?

II. Le nostre scelte quotidiane – cosa acquistare, dove cenare, cosa pubblicare sui social – sono sempre più **influenzate da applicazioni basate sull'intelligenza artificiale**. Queste applicazioni non sono neutrali: fanno parte di **architetture digitali della scelta**, cioè ambienti progettati intenzionalmente per **guidare il comportamento degli utenti**, spesso sfruttando **bias cognitivi** noti. Se il contesto digitale orienta le nostre decisioni senza che ce ne rendiamo conto, possiamo ancora parlare di **scelte libere e autonome**? E soprattutto: questo condiziona la nostra **identità**?

Etica in Silicio: I Robot Possono Essere Morali?

III. Robot come **agenti morali**: quale idea di etica è in gioco?

I. INTELLIGENZA ARTIFICIALE E PROCESSI DECISIONALI: IL FUTURO DECISORE È UN SOFTWARE?

- **L'IA generativa: compiti e attività** che richiedevano riflessione e discernimento e che erano dominate dal giudizio umano vengono sempre più **affidati a sistemi automatizzati**.
- 1) In ambito giudiziario: **COMPAS** (Correctional Offender Management Profiling for Alternative Sanctions): un sistema di **valutazione del rischio algoritmico** sviluppato da Northpointe (oggi parte di Equivant). Viene utilizzato dai tribunali statunitensi per stimare il rischio di (a) recidiva; (b) mancata comparizione in tribunale; (c) di comportamenti violenti. I giudici usano il punteggio di COMPASS per decidere se negare o concedere la libertà su cauzione;
 - 2) In ambito medico: (a) software per la lettura delle immagini diagnostiche; (b) per **sviluppare vaccini** (Moderna (mRNA-1273 – ha progettato l'RNA che codifica la proteina SPIKE); (c) aiuta a prendere velocemente le **decisioni di TRIAGE** o per organizzare le risorse in casi di disastri naturali.

I. INTELLIGENZA ARTIFICIALE E PROCESSI DECISIONALI: IL FUTURO DECISORE È UN SOFTWARE?

- I **sistemi di decisione automatica** (reti neurali artificiali) **sono black boxes**: in molti casi non abbiamo un'idea di come e perché una scelta (output) sia stata ottenuta dai dati di partenza (input).
 - Un **modello di calcolo (es.)** costituito da un *sistema di connessione* fra *nodi* organizzato a strati tale che (a) ciascun nodo è connesso a tutti i nodi dello strato successivo e (b) ogni connessione ha un determinato valore.
 - Il **processo decisionale** della macchina **è controllabile e manipolabile** dal ricercatore. Attraverso la variazione di fattori (input, nodi, connessioni, ecc.) modifico il processo che connette input ad output. Tale processo è **descrivibile attraverso funzioni matematiche complesse**.
 - Il suo **percorso di calcolo** spesso **non è descrivibile attraverso il linguaggio naturale**: una descrizione che **spiega attraverso ipotesi e inferenze razionali come un determinato input è associato ad un determinato output**. Il comportamento del sistema è **razionalmente opaco o semanticamente inaccessibile**.

I. INTELLIGENZA ARTIFICIALE E PROCESSI DECISIONALI: IL FUTURO DECISORE È UN SOFTWARE?

ES: Sistema di decisione automatica impiegato per fornire **stime sulla probabilità di contrarre una malattia** grave (S. ZOPPI CAIANI 2024)

1. **Addestramento:** il sistema ha **accesso ad un database** di immagini esaminate da esperti e adegua – tramite un algoritmo - il proprio comportamento alle stime dei medici.
2. Il sistema, tuttavia, **impara** ricorrendo ad una **computazione molto più complessa** di quella adottata dai medici che **per ogni immagine considera migliaia di variabili** il cui valore è soppesato alla luce di altre variabili che lo porta a ‘scovare’ **modelli ricorrenti** (es: una certa sfumatura di grigio è presente in pazienti che sviluppano una malattia)
3. Il sistema può manifestare un **livello di accuratezza diagnostica maggiore** del medico esperto (la possibilità di sviluppare una malattia è correlata con un numero maggiore di variabili).
4. Le **correlazioni statistiche** (cioè i legami tra certi dati iniziali e la probabile malattia) trovate dall’IA possono sorprendere i medici: non riescono a **spiegarle biologicamente** perché quei dati siano collegati causalmente alla malattia.
5. Queste correlazioni **non hanno interpretazione semantica**: cioè, non possiamo attribuire loro un significato biologico chiaro (non sappiamo perché funzionano).
6. Senza un’**interpretazione umana** o un modello biologico esplicito, **non possiamo costruire una teoria medica** su queste correlazioni: abbiamo un **sistema efficace** (ha buone performance predittive!), ma è “**opaco**”.

I. EFFICIENZA E AFFIDABILITÀ. IMPLICAZIONI ETICHE

1. Responsabilità e affidabilità

- Se non capiamo come l'IA prende una decisione, **chi è responsabile** di un eventuale errore? (a) Il medico? (b) Il programmatore? (c) Il sistema stesso?
- In medicina, decidere su diagnosi o trattamenti **richiede trasparenza**, perché in gioco c'è la **vita e la salute delle persone**.

2. Fiducia e consenso informato

- **Il paziente ha diritto a conoscere i criteri su cui si basa una diagnosi.**

Se questa proviene da un sistema che “**non possiamo spiegare**”, si apre una questione importante: (a) si rischia di **compromettere il consenso informato e la fiducia nel rapporto medico-paziente**, oppure (b) questo limite **non ha un impatto** significativo su questi aspetti?

Art. 1

Consenso informato

1. La presente legge, nel rispetto dei principi di cui agli articoli 2, 13 e 32 della Costituzione e degli articoli 1, 2 e 3 della Carta dei diritti fondamentali dell'Unione europea, tutela il **diritto** alla vita, alla salute, **alla dignità** e **all'autodeterminazione della persona** e stabilisce che nessun trattamento sanitario può essere iniziato o proseguito **se privo del consenso libero e informato della persona** interessata, tranne che nei casi espressamente previsti dalla legge.
2. **È promossa e valorizzata la relazione di cura e di fiducia** tra paziente e medico che si basa sul consenso informato nel quale si incontrano l'autonomia decisionale del paziente e la competenza, l'autonomia professionale e la responsabilità del medico. Contribuiscono alla relazione di cura, in base alle rispettive competenze, gli esercenti una professione sanitaria che compongono l'equipe sanitaria. In tale relazione sono coinvolti, se il paziente lo desidera, anche i suoi familiari o la parte dell'unione civile o il convivente ovvero una persona di fiducia del paziente medesimo.
3. Ogni persona ha **il diritto** di conoscere le proprie condizioni di salute e **di essere informata in modo completo, aggiornato e a lei comprensibile riguardo alla diagnosi, alla prognosi, ai benefici e ai rischi** degli accertamenti diagnostici e **dei trattamenti sanitari indicati**, nonché riguardo alle possibili alternative e alle **conseguenze dell'eventuale rifiuto** del trattamento sanitario e dell'accertamento diagnostico o della rinuncia ai medesimi. Può rifiutare in tutto o in parte di ricevere le informazioni ovvero indicare i familiari o una persona di sua fiducia incaricati di riceverle e di esprimere il consenso in sua vece se il paziente lo vuole. Il rifiuto o la rinuncia alle informazioni e l'eventuale indicazione di un incaricato sono registrati nella cartella clinica e nel fascicolo sanitario elettronico.[...]
6. **Il medico è tenuto a rispettare la volontà espressa dal paziente di rifiutare il trattamento sanitario o di rinunciare al medesimo** e, in conseguenza di ciò, è esente da responsabilità civile o penale. Il paziente non può esigere trattamenti sanitari contrari a norme di legge, alla deontologia professionale o alle buone pratiche clinico-assistenziali; a fronte di tali richieste, il medico non ha obblighi professionali.

II. AMBIENTI DIGITALI DELLA SCELTA, *NUDGES* DIGITALI E VALORI MORALI

- **La scelta** del luogo dove cenare, il monitoraggio dei parametri corporei e medici, l'acquisto di beni e servizi sono attività **mediate da applicazioni e social network** che si fondano su **tecnologie** basate sull'IA note come «**nudges digitali**» (spinte gentili)
- I nudges digitali, all'interno di particolari ambienti digitali di scelta, **sfruttano** a loro vantaggio gli automatismi e le «**euristiche**» umane **in modo non trasparente**
- Le euristiche sono **scorciatoie cognitive** utilizzate tipicamente dagli esseri umani per risolvere problemi pratici complessi quando manca il tempo per prendere decisioni meditate. Sebbene siano utili, producono spesso errori e distorsioni cognitive(TVERSKY e KAHNEMAN 1972), i cosiddetti *bias*.
 - Es: *bias della disponibilità*; *bias dello status quo*

II. AMBIENTI DIGITALI DELLA SCELTA, *NUDGES DIGITALI E VALORI MORALI*

- **Bias della disponibilità:** errore mentale che ci porta a stimare la probabilità o rilevanza di un evento **in base a quanto lo ricordiamo facilmente**, non a dati reali. È una scorciatoia utile, ma può farci **sopravvalutare** eventi rari e **sottovalutare** quelli comuni, generando **decisioni poco razionali**.
Accade **soprattutto con eventi recenti, emotivi o amplificati dai media**. Utile, sì, ma spesso ingannevole.

- Alcuni effetti:

1. Giudizi morali distorti

Diamo più peso a **problemi mediatici** (es. terrorismo) e trascuriamo emergenze gravi ma meno visibili (es. fame, malattie). Un'etica razionale (come l'utilitarismo) invita a valutare le **conseguenze reali**, non le emozioni.

2. Empatia selettiva

Siamo più toccati da **casi individuali** noti che da milioni di vittime invisibili. È la "**fallacia della vittima identificabile**".

3. Decisioni pubbliche ed etica delle priorità

Governanti e cittadini possono **preferire investire in ciò che fa notizia**, ignorando problemi cronici più gravi (priorità sanitarie, scuola, ecc.). Eticamente, questo apre un dibattito su **equità e uso razionale delle risorse**.

II. AMBIENTI DIGITALI DELLA SCELTA, NUDGES DIGITALI E VALORI MORALI

- **Bias dello status quo:** Tendiamo a preferire ciò che conosciamo, anche se il cambiamento potrebbe portarci benefici maggiori. Perché è un bias? Perché familiarità ≠ bontà. Attribuiamo valore emotivo alla stabilità, anche se non è la scelta più razionale.
- È una **inerzia mentale:**
 - Valorizziamo troppo la situazione attuale;
 - Sottovalutiamo i vantaggi del cambiamento;
 - Temiamo di perdere più di quanto speriamo di guadagnare (→ *loss aversion*);
- **Alcuni effetti:**
 1. **Piani pensionistici:** si resta nel default anche se esistono alternative migliori; **scelte mediche:** si rifiuta una cura nuova per restare nella “sicurezza” del noto; **abitudini digitali:** non si cambia software o dieta per pura consuetudine;
 2. Ostacola il **progresso morale** (es. diritti civili), cioè contro le etiche riformiste (es. Singer): lo status quo va valutato **come qualsiasi altra opzione**, non come punto di partenza “neutro”.

II. AMBIENTI DIGITALI DELLA SCELTA, NUDGES DIGITALI E VALORI MORALI

- I «nudges digitali» coinvolgono molteplici valori morali. La manipolazione attraverso nudges digitali viola l' **autonomia**, la **libertà** e l' **identità**.
1. **Autonomia**: una persona è autonoma se (a) possiede ed esercita alcune competenze (cognitive, psicologiche, sociali, emotive) e (b) può compiere scelte con impegni che sottoscrive (**autenticità**).
I nudges digitali impediscono di esercitare in modo adeguato i processi razionali e riflessivi, **ampliando stati emotivi; instillando credenze false, sfruttando le debolezze personali** (SUSTEIN 2021). Il problema principale è la mancanza di **trasparenza** (BOTES 2023)
 2. **Libertà**: il problema non è tanto la presenza di interferenze (violazione della libertà negativa, BERLIN) ma la presenza **del dominio arbitrario altrui** (assenza di libertà «repubblicana», SKINNER, PETTIT).
I nudges digitali violano la libertà come assenza di dominio: **le nostre capacità decisionali dipendono** continuamente dagli **aspetti della realtà** che sono mostrati nelle strutture digitali in cui ci muoviamo (GALLETTI 2024)

II. AMBIENTI DIGITALI DELLA SCELTA, *NUDGES* DIGITALI E VALORI MORALI

3. Identità:

- I. L'identità come **tutela della privacy** dell'individuo;
- II. Come **individualità** contrapposta a **omologazione**. Gli ambienti digitali confinano le persone in 'bolle di filtraggio' impediscono loro di essere esposte ad una varietà di opinioni. La loro identità personale si trasforma in un **prodotto** costruito attorno ad una **categoria generalizzante** che crea nell'individuo aspettative, interessi e desideri. Inoltre: gli schemi cognitivi e le assunzioni valoriali che prefigurano i campi di scelta **sono nascosti**. L'esperienza con gli altri è costruita attorno ad un **modello reputazionale**.

Queste tecnologie sembrano realizzare un impoverimento del **progetto umanistico** elaborato nell'etica moderna (Hume; J. S. Mill e Dewey) che ha a cuore **lo sviluppo libero e creativo della personalità individuale** da parte del maggior numero di individui.

Il progetto umanistico richiede è che le persone possano fare esperienza del mondo e degli altri e che questa esperienza sia caratterizzata **dall'incontro con qualcosa di diverso da sé**, qualcosa di solido che il proprio sé incontra e che gli resiste e **dal cui punto di vista l'individuo può vedersi e comprendersi**.

ETICA E INTELLIGENZA ARTIFICIALE: PRINCIPI FONDAMENTALI

- Le tecnologie intelligenti sollevano questioni etiche profonde, che richiedono di applicare concetti normativi. Ci chiediamo, ad esempio:
 - In che modo queste tecnologie **utilizzano o trasformano concetti morali fondamentali**, come l'autonomia e la responsabilità?
 - **Quali criteri dovrebbero guidare progettisti e autorità pubbliche** nello sviluppo e nella regolazione di questi sistemi?
- Molti dei **principi emersi in questo dibattito** non sono nuovi: provengono dalla riflessione bioetica e si ritrovano anche in documenti ufficiali come le **“Ethics Guidelines for Trustworthy AI”** (pubblicate **9 aprile 2019** dalla **Commissione Europea**, tramite il gruppo di esperti indipendenti chiamato **AI HLEG** - High-Level Expert Group on Artificial Intelligence).

I QUATTRO PRINCIPI ETICI DELL'UNIONE EUROPEA PER UN'IA AFFIDABILE

Secondo il documento, per garantire che l'IA sia sviluppata e usata in modo etico, bisogna rispettare **quattro principi fondamentali**, radicati nella **Carta dei diritti fondamentali dell'UE**. Questi principi sono considerati «imperativi etici» per tutti gli attori coinvolti nel settore.

1. Rispetto dell'autonomia umana

- Le persone devono poter **controllare i sistemi di IA**, senza esserne **manipolate** o escluse. L'IA dovrebbe **rafforzare**, e non ridurre, le **capacità cognitive**, sociali e culturali dell'essere umano, sia a livello individuale che collettivo.

2. Prevenzione dei danni

- L'IA non deve compromettere la **dignità umana**, l'**integrità fisica e mentale** né l'**ambiente**. Bisogna prestare attenzione anche alle **asimmetrie di potere**, evitando che l'IA le accentui.

3. Equità

- Richiede una **distribuzione giusta** di costi e benefici, **assenza di discriminazioni** e **pari opportunità**. Include anche il **diritto a contestare decisioni** prese da sistemi automatizzati.

4. Esplicabilità

- Le decisioni dell'IA devono essere **trasparenti**, **comprendibili** e **comunicabili in modo chiaro**, così che possano essere valutate e, se necessario, **messe in discussione**.

EU AI ACT (ARTIFICIAL INTELLIGENCE ACT)

- **Proposta nel 2021**, approvata dal Parlamento Europeo a marzo **2024**. Prima legge al mondo per **regolare l'IA in modo orizzontale**
- Si applica a **tutti i contesti** in cui l'IA viene usata: Sanità; Giustizia; Educazione; Sicurezza pubblica; Lavoro; Marketing; Servizi pubblici, Ecc.
- Classifica i sistemi in **rischio inaccettabile, alto, limitato, minimo** (valutazione del rischio prima di essere immessa sul mercato e possibile divieto di produzione);
- Mira a garantire **sicurezza, trasparenza e diritti fondamentali (contro la manipolazione subliminale; immagini sul lavoro, ecc.)**

III. UN'ETICA PER I ROBOT E INTELLIGENZA ARTIFICIALE

- È **importante** progettare **tecnologie autonome** (come robot e IA) che **agiscano secondo criteri etici**, anche in assenza di un intervento umano diretto.
- IA dipende dai **dati e algoritmi** usati per addestrarla. Se i dati riflettono **pregiudizi umani** (etnia, genere, status...), l'IA li **replica su larga scala**, spesso in modo invisibile, perché viene percepita come neutrale.

III. UN'ETICA PER I ROBOT E INTELLIGENZA ARTIFICIALE

Inoltre ... desideriamo che gli agenti artificiali siano:

- **Favorevoli all'umanità in generale**, senza perseguire scopi propri contrari all'interesse umano (come nelle opere di fantascienza).
 - Es: **Eticamente responsabili** secondo principi condivisi, come quelli della **conferenza di Asilomar** (2017) o le **Leggi della robotica** di Asimov.
- **Rischi esistenziali e intelligenza artificiale.**
Secondo l'approccio di Nick Bostrom, tra i **rischi globali** che minacciano l'intera umanità c'è anche quello legato allo sviluppo di un'**IA non amichevole**, potenzialmente capace di annientarla.



NICK BOSTROM
SUPERINTELLIGENZA
TENDENZE, PERICOLI, STRATEGIE



Bollati Boringhieri

III. UN'ETICA PER I ROBOT E INTELLIGENZA ARTIFICIALE

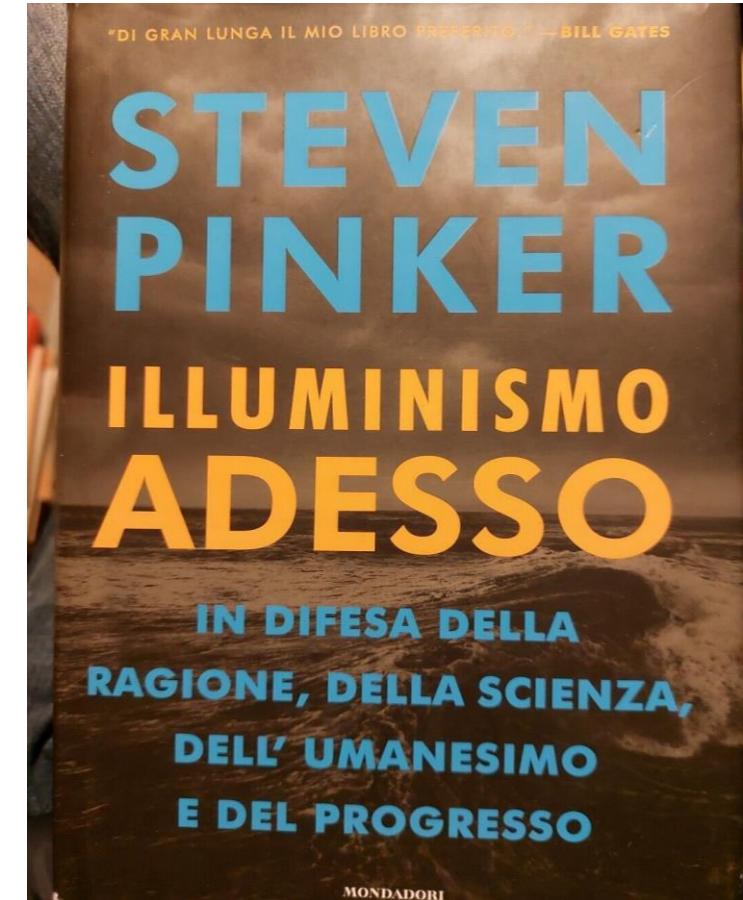
- **Critica di Steven Pinker al concetto di IA non amichevole**

➤ l'IA è solo un insieme di **tecniche progettate per scopi specifici**, basate su dati limitati e contestuali. Non si tratta di intelligenza onnisciente: opera sempre all'interno di **problemi concreti e definiti**.

➤ Inoltre, anche un'eventuale IA superintelligente sarebbe **vincolata alla realtà fisica**, necessitando di:

- Hardware
- Energia
- Produzione industriale

- In sintesi, **non può esistere una superintelligenza onnipotente svincolata dalla realtà materiale** del pianeta e dell'universo.



III. UN'ETICA PER I ROBOT E INTELLIGENZA ARTIFICIALE

- Ci sono due tipi di agenti morali artificiali:

1) Agenti morali impliciti

- Sono progettati per **agire in modo eticamente accettabile**, ma **senza comprendere o rappresentare** concetti morali.

Seguono **regole preimpostate** o comportamenti appresi dai dati, ma **non possono spiegare** le loro azioni in termini morali.

Esempio: un'auto autonoma che frena davanti a un pedone perché è programmata per farlo, non perché "sa" che è giusto.

2) Agenti morali esplicativi

- Sono sistemi dotati di **modelli interni dell'etica** e capaci di **ragionare su dilemmi morali**. Possono valutare alternative, considerare valori, regole o conseguenze, e in alcuni casi **spiegare le loro scelte**.

Esempio: un robot assistente valuta se violare la privacy per aiutare una persona in pericolo, decidendo secondo un principio morale (es. utilitarismo o deontologia).

DILEMMI DEL CARRELLO

- **Scenario principale – Il dilemma del carrello**
- Un carrello è fuori controllo su un binario.
Davanti a sé: **5 persone legate**.
Accanto a te: una leva che può deviarlo su un binario secondario, dove c'è **1 sola persona**.
- **Domanda:**
E' lecito **deviare** il carrello, uccidendo una persona per salvarne cinque?
- **Variante – L'uomo grasso**
- Stesso carrello, stessa situazione.
Sei su un ponte sopra i binari.
Accanto a te: un uomo molto corpulento.
Spingerlo giù fermerebbe il carrello, salvando i 5.
- **Domanda:**
È lecito **spingere** un uomo per salvare gli altri?

III. UN'ETICA PER I ROBOT E INTELLIGENZA ARTIFICIALE

- **In condizioni estreme** (Katrina a N.O.), gli **agenti artificiali morali esplicativi** sono la **migliore** possibilità che abbiamo di **operare eticamente** (J. H. Moore)
- **Quale immagine dell'etica? Etica procedurale:** il ragionamento etico è una forma di procedura decisionale che imita il diritto nei suoi aspetti di chiusura del sistema normativo. L'etica procedurale è **importante** per macchine dotate di grande autonomia (robot militari).
- **La dimensione dell'etica non è solo questa!** Mill sosteneva che il fondamento ultimo della morale è da cercare nei **sentimenti di unione solidale** con gli altri umani, e che ci legano anche agli altri esseri senzienti. Questa dimensione comporta «**indeterminatezza**» e »**discrezionalità**». Le situazioni problematiche sono in grado di **educarci**.
- La **questione** sollevata dalle «macchine morali» **è quella della riorganizzazione della società con l'attribuzione di attività importanti** ad agenti che sono **sottratti alla più ampia sfera della morale**. Vogliamo davvero ridurre le basi del nostro interesse umano e morale?
 - Es: **robot militari?** Il soldato cittadino democratico non è un mero esecutore delle regole di chi ha il comando. Due dimensioni della riflessione del soldato: cittadinanza e umanità;
 - Es: **robot per la cura degli anziani.** La cura è espressione di attenzione e affetto. L'assistenza robotica affettiva è una forma di inganno e manipolazione.
 - Es: **amministrazione della giustizia.** Il giudice interpreta le norme, non le applica meccanicamente.