Politecnico di Torino

Nexa Center
*for Internet & Society*

*Per il ciclo di incontri "i Nexa Lunch Seminar" (ogni 4° mercoledì del mese)*

106° Nexa Lunch Seminar

# L'«etica» come specchietto per le allodole
# Sistemi di intelligenza artificiale
# e violazioni dei diritti


**DANIELA TAFANI**

(Università di Pisa)


**Mercoledì 24 maggio 2023, ore 13.00**
(termine: ore 14.00)

# Intelligenza artificiale e narrazioni

«Intelligenza artificiale», oggi:

1. **Una tecnologia**, entro uno spettro che va dai sistemi esperti, fino, teoricamente, all'intelligenza sintetica. Al momento, per arrivare a quest'ultima, non abbiamo nemmeno le domande giuste.

2. **Un atto linguistico**, una narrazione che deriva in gran parte dalla fantascienza su computer e robots con capacità straordinarie e tendenze apocalittiche, un discorso che evoca una precisa percezione sociale dell'IA come tecnologia. Un'espressione di marketing.

   E' utilizzata per esercitare controllo, o assicurarsi profitto, senza assumersi responsabilità.
   Non mera evasione dalla responsabilità: creazione di asimmetrie di potere.

Peter R. Lewis, Stephen Marsh, Jeremy Pitt, *AI vs «AI»: Synthetic Minds or Speech Acts,* «IEEE Technology and Society Magazine», 2021, https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9445758

# Intelligenza artificiale: di cosa stiamo parlando

1. **Intelligenza artificiale simbolica:** così come l'informatica, deriva dalla logica, ossia dalla disciplina che studia il ragionamento deduttivo**.**

   I sistemi simbolici sono sempre in grado di spiegare perché hanno fornito una determinata soluzione o hanno preso una qualsiasi «decisione».

2. **Intelligenza artificiale sub-simbolica:** è di natura statistica. Il focus si è spostato dai linguaggi di programmazione usati per «istruire» la macchina ai dati usati per permettere alla macchina di «imparare» in modo autonomo

   I sistemi sub-simbolici, al momento, non hanno la capacità di fornire spiegazioni per i risultati prodotti.

Maurizio Gabbrielli, *L'intelligenza artificiale cos'è?*, in *XXVI Lezioni di diritto dell'intelligenza artificiale*, a cura di Ugo Ruffolo, Torino, Giappichelli, 2021.

# IA ristretta e IA generale

**Narrow AI (o weak AI)** i sistemi di IA ristretta sono sistemi che possono eseguire uno o pochi compiti specifici.

**General AI (o strong AI):** un sistema di IA generale è un sistema che può eseguire la maggior parte delle azioni che gli esseri umani possono compiere.

Tutti i sistemi di IA attualmente in uso sono esempi di IA ristretta: funzionano per i compiti particolari per i quali sono stati programmati, a condizione che ciò che incontrano non sia troppo diverso da quello che hanno sperimentato in precedenza.

G. Marcus, E. Davis, *Rebooting AI. Building Artificial Intelligence We Can Trust*, New York, Pantheon Books, 2019.

# Sistemi di apprendimento automatico (*machine learning*)

Nella famiglia di tecnologie denominata «intelligenza artificiale» – che si occupa di realizzare strumenti (software e hardware) che siano capaci di eseguire compiti normalmente associati all'intelligenza naturale – l'apprendimento automatico ha reso possibile, per alcuni specifici compiti non trattabili con l'intelligenza artificiale simbolica, un rapido e genuino progresso:

- previsione e generazione di stringhe di testo,
- riconoscimento facciale,
- ricerca per immagini,
- identificazione di contenuti musicali.

I sistemi di apprendimento automatico, di natura sostanzialmente statistica, consentono di costruire modelli a partire da esempi, in un processo iterativo di minimizzazione della distanza rispetto ai risultati attesi, purché si abbiano a disposizione

- potenti infrastrutture computazionali,
- enormi quantità di dati.

Le grandi aziende tecnologiche che, intorno al 2010, in virtù di un modello di business fondato sulla sorveglianza, detenevano già l'accesso al mercato necessario per l'intercettazione di grandi flussi di dati e metadati individuali e le infrastrutture di calcolo per la raccolta e l'elaborazione di tali dati, hanno potuto perciò raggiungere, con l'applicazione di algoritmi in gran parte già noti da decenni, traguardi sorprendenti.

Le grandi aziende tecnologiche hanno colto l'opportunità di un'espansione illimitata di prodotti e servizi «intelligenti»:

se un sistema di «intelligenza artificiale» è in grado di tradurre quello che scriviamo, perché non sostenere che sia anche in grado di comprenderlo?

Se può identificare un singolo individuo o classificarne correttamente alcuni tratti somatici, perché non sostenere che sia in grado altresì di riconoscere un ladro o un bravo lavoratore dalle loro fattezze esteriori o un malato di mente dalla voce ?

Perché non trasformare un sistema statistico, grazie alla polvere magica dell'«intelligenza artificiale», in un oracolo in grado di prevedere i futuri reati di ogni individuo o la «futura performance» scolastica di ogni singolo studente?

# Prodotti pericolosi e non funzionanti

Con la formula di marketing «intelligenza artificiale», hanno diffuso e messo in commercio sistemi di apprendimento automatico per lo svolgimento di attività che tali sistemi non sono in grado di svolgere o che semplicemente non sono possibili.

Tra i prodotti di questo genere ci sono

1. le auto a guida autonoma,
2. i sistemi di ottimizzazione predittiva,
3. i generatori di linguaggio naturale.

# Le auto a guida autonoma e il dilemma del trolley

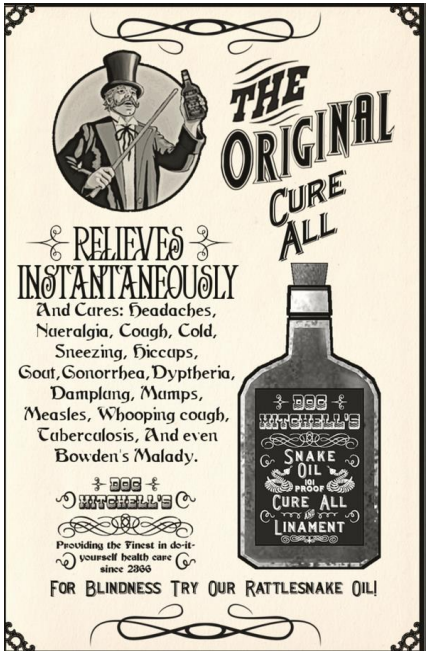
Allo stato attuale della tecnologia, le auto «a guida autonoma» non sono a guida autonoma.

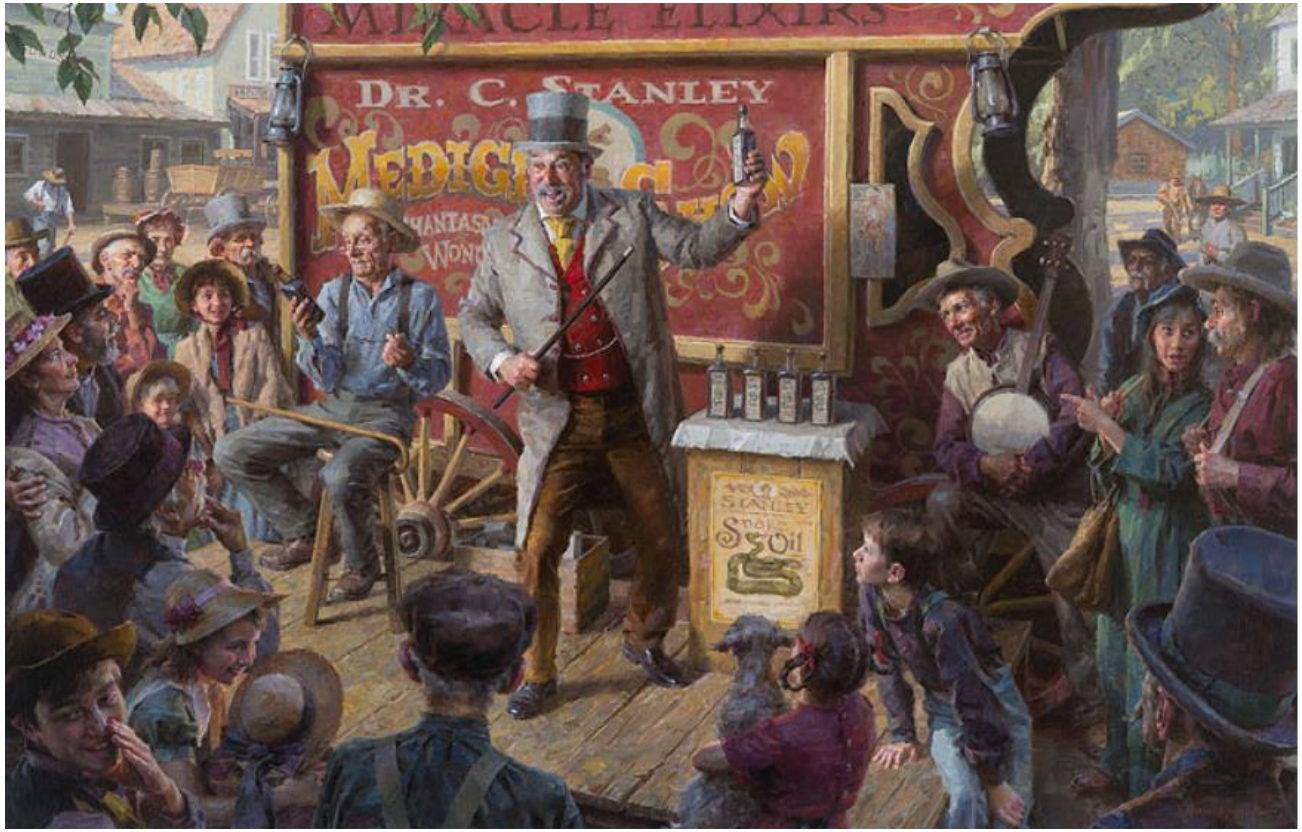
L'idea che, in caso di incidente, la tecnologia delle auto a guida autonoma ponga dilemmi etici straordinari - come le scelte che le auto stesse dovrebbero compiere - piuttosto che le consuete questioni di sicurezza, trasparenza, controllo e cautela, nasce dalla considerazione antropomorfa di tali veicoli come agenti morali.

Narrazione generata, a fini di marketing, dai produttori.
Operazione consueta, nel campo dell'etica dell'IA: si introducono nel mercato filosofico casi moralmente difficili, con i quali gli specialisti e il grande pubblico possano baloccarsi per qualche anno, distraendo così l'opinione pubblica e le istituzioni dal fatto che i sistemi di ML semplicemente non funzionano, e quindi causano danni, quando siano utilizzati per compiti che, come la guida sulle comuni strade urbane, richiederebbero l'intelligenza artificiale generale.

H. Jin, *Tesla video promoting self-driving was staged, engineer testifies*, «Reuters», January 18, 2023, https://www.reuters.com/article/tesla-autopilot-video-idTRNIKBN2TW1EB ; P. Valdes-Dapena, *Tesla: Our 'failure' to make actual self-driving cars 'is not fraud'*, December 12, 2022, https://edition.cnn.com/2022/12/12/business/tesla-fsd-autopilot-lawsuit/index.html

# I sistemi di ottimizzazione predittiva




https://aisnakeoil.substack.com/

# Against Predictive Optimization:

## On the Legitimacy of Decision-Making Algorithms that Optimize Predictive Accuracy

Angelina Wang, Sayash Kapoor, Solon Barocas, Arvind Narayanan.

FAccT 2023

**Read our draft paper**

## Our argument

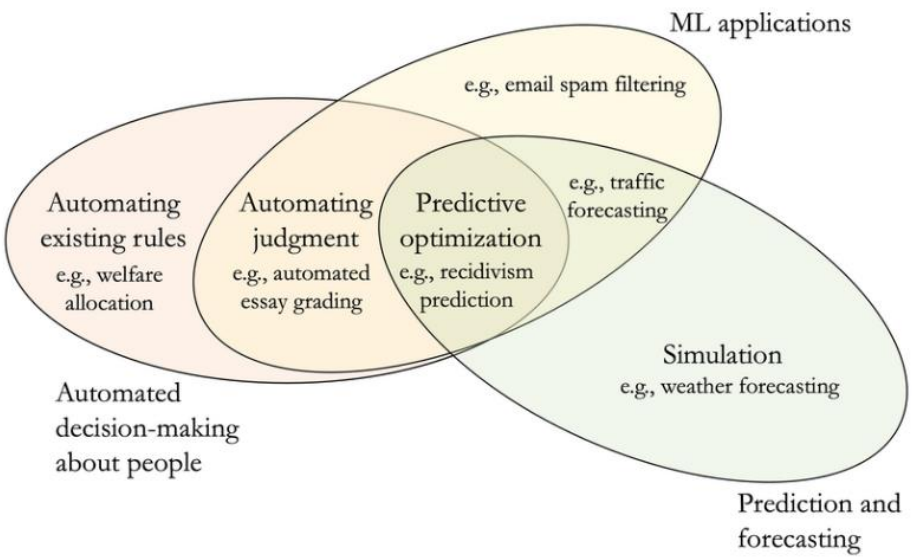
Predictive optimization is a distinct type of automated decision making that has **proliferated widely**. It is sold as accurate, fair, and efficient.

We identify a **recurring set of flaws** that apply broadly to predictive optimization, are hard to fix technologically, and negate its claimed benefits.

Any application of predictive optimization should be considered **illegitimate** by default unless the developer justifies how it avoids these flaws.

## What is predictive optimization?

We coin the term predictive optimization to refer to automated decision-making systems where machine learning is used to make predictions about some future outcome pertaining to individuals, and those predictions are used to make decisions about them.




https://predictive-optimization.cs.princeton.edu/

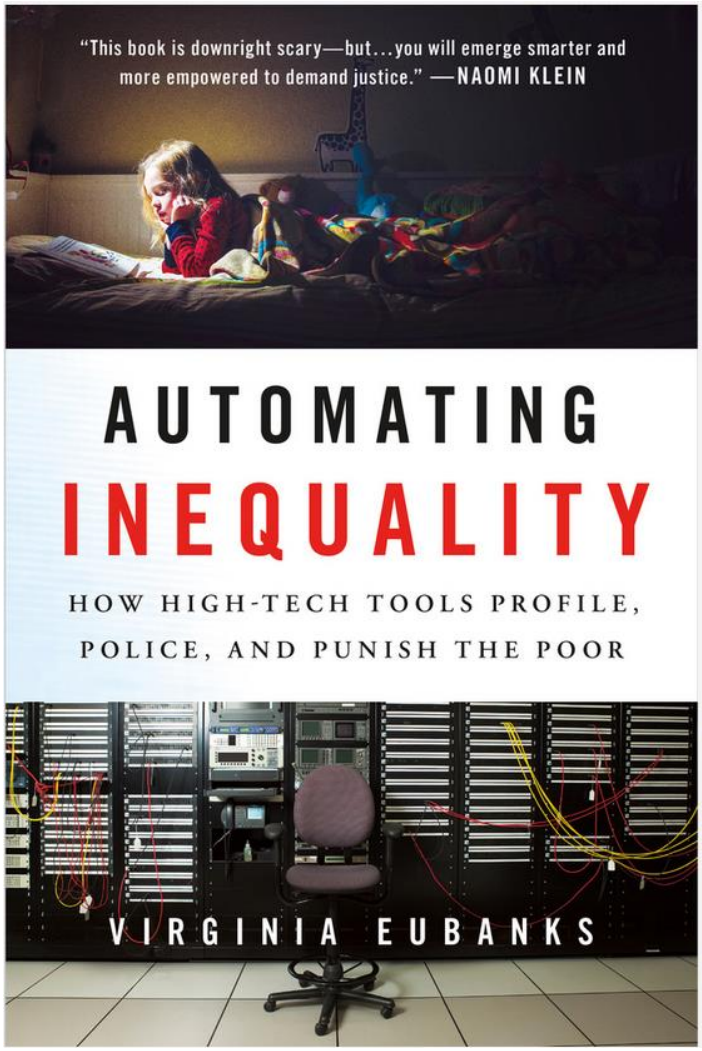
**Non possiamo predire il futuro comportamento di singoli esseri umani. Nemmeno utilizzando l'IA.**

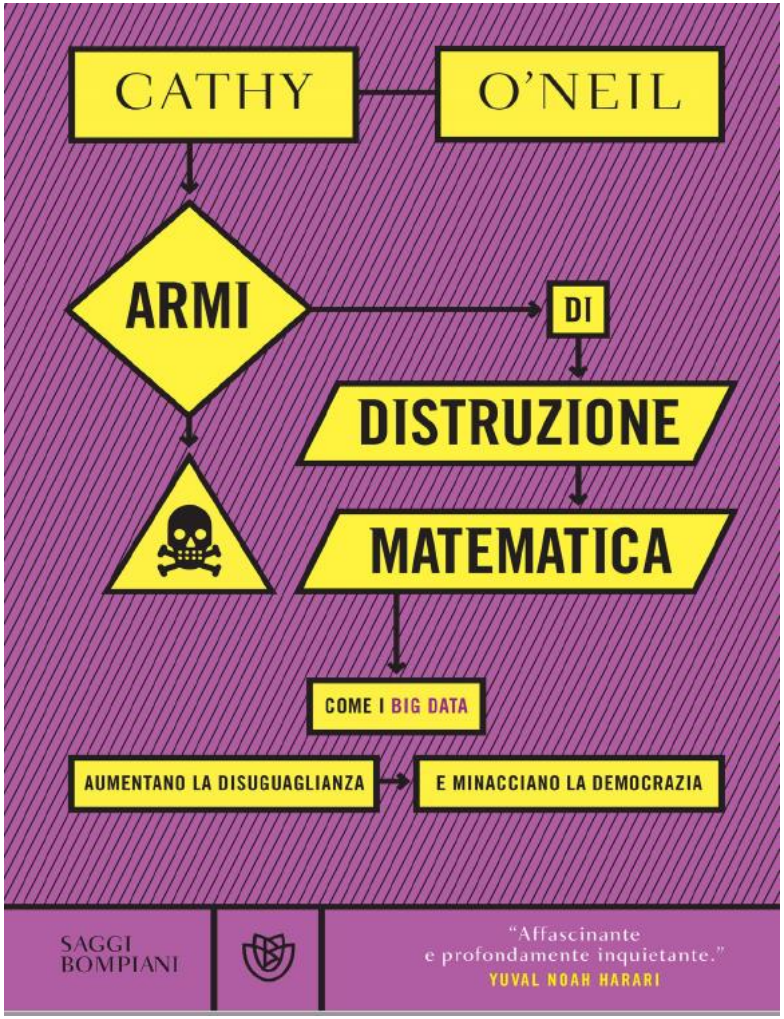
Per la previsione dei risultati sociali, l'IA non è significativamente migliore di un calcolo manuale che utilizzi solo alcune caratteristiche.

Le applicazioni di IA che promettono di predire esiti sociali sono fondate su pseudoscienze, sono «olio di serpente».

A una verifica puntuale, tali sistemi risultano così poco affidabili, per la previsione di eventi e azioni individuali, che alcuni ricercatori suggeriscono piuttosto il ricorso a una lotteria tra le persone ammissibili.

A. Wang, S. Kapoor, S. Barocas, A. Narayanan, Against Predictive Optimization: On the Legitimacy of Decision-Making Algorithms that Optimize Predictive Accuracy, October 4, 2022, pp. 1-29, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4238015


"This book is downright scary—but...you will emerge smarter and more empowered to demand justice." —NAOMI KLEIN
AUTOMATING
INEQUALITY
HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR
VIRGINIA EUBANKS


CATHY
O'NEIL
ARMI
DI
DISTRUZIONE
MATEMATICA
COME I BIG DATA
AUMENTANO LA DISUGUAGLIANZA
E MINACCIANO LA DEMOCRAZIA
SAGGI BOMPIANI
"Affascinante e profondamente inquietante."
YUVAL NOAH HARARI

# L'«etica dell'intelligenza artificiale»

Giudizi e decisioni che hanno effetti rilevanti sulle vite di esseri umani sono oggi affidati, in un numero crescente di ambiti, a sistemi di intelligenza artificiale che non funzionano.

Tali malfunzionamenti non sono occasionali e non sono scongiurabili con interventi tecnici: essi rivelano, anzi, il funzionamento ordinario dei sistemi di apprendimento automatico.

Considerato il ruolo cruciale di tali sistemi nel modello di business delle grandi aziende tecnologiche, esse mirano a sottrarre tali prodotti all'intervento giuridico: è nata così, come operazione di cattura culturale, con l'obiettivo di rendere plausibile un regime di mera autoregolazione, l'«etica dell'intelligenza artificiale» .

# A cosa serve l'«etica dell'intelligenza artificiale»

Tutela e legittimazione del modello di business delle grandi aziende tecnologiche.

L'impostazione del discorso è dettata dalla sua funzione.
Perciò:

- determinismo tecnologico: assunto dell'inevitabilità e «logica del fatto compiuto»;
- soluzionismo: questioni sociali trattate quali problemi di disciplinamento e controllo, passibili di soluzioni tecniche;
- set di problemi e soluzioni che individua nel *design* tecnico il livello appropriato per la soluzione di tutti i problemi e pone al riparo dalla discussione le decisioni di business;
- tema dei *bad actors*, anziché dei rapporti di sfruttamento o di potere strutturali.

D. Greene, A.L. Hoffman, L. Stark, *Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning*, in *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019, pp. 2122-2131, http://hdl.handle.net/10125/59651

# Sistemi di apprendimento automatico

1. Dati
2. Potenza di calcolo
3. Algoritmi

Altre «estrazioni», non tematizzata dai discorsi sull'etica:

4. Terre rare
5. Energia
6. Lavoro

Kate Crawford, *Né intelligente né artificiale. Il lato oscuro dell'IA*, Il Mulino, Bologna 2021

# Terre rare

1555, Georg Agricola: è chiaro che il danno che provoca l'estrazione dei minerali è superiore al valore dei metalli che essa produce.

Logica dell'estrazione: «costante prelievo di minerali, acqua e combustibili fossili, sostenuto dalla violenza delle guerre, dell'inquinamento, dell'estinzione e dell'esaurimento delle risorse».

Occorre riflettere sui media come processi anche geologici: esaurimento di risorse non rinnovabili.

Per raffinare una tonnellata di terre rare, il processo produce 75.000 litri di acqua acida e una tonnellata di residui radioattivi.

Es. Tesla utilizza più di 28mila tonnellate l'anno di idrossido di litio, la metà del consumo totale del pianeta (circa 60 kg di litio per ogni auto elettrica).

Kate Crawford, *Né intelligente né artificiale. Il lato oscuro dell'IA*, Il Mulino, Bologna 2021.

# Energia

Mito della tecnologia pulita.
In realtà, già oggi l'emissione di anidride carbonica da parte delle infrastrutture computazionali è pari a quella dell'industria aeronautica.

I data center sono tra i maggiori consumatori di elettricità al mondo.

Es. l'esecuzione di un solo modello di elaborazione del linguaggio naturale ha prodotto 300.000 chilogrammi di anidride carbonica (quanto 5 auto a gas nel loro intero ciclo di vita, produzione compresa, o 125 voli andata e ritorno da New York a Pechino).

Kate Crawford, *Né intelligente né artificiale. Il lato oscuro dell'IA*, Il Mulino, Bologna 2021.
S. Luccioni, *The mounting human and environmental costs of generative AI*, April 12, 2023, https://arstechnica.com/gadgets/2023/04/generative-ai-is-cool-but-lets-not-forget-its-human-and-environmental-costs/

# Lavoro: l'intelligenza artificiale «fatta a mano»

Il numero dei lavoratori pagati per realizzare interamente, o semplicemente accompagnare, il lavoro delle IA, svela che siamo in realtà agli antipodi delle fantasie robotiche che alimentano l'immaginario degli investitori e dei media: qui vediamo soltanto una miriade di proletari del clic, lavoratori non specializzati che svolgono le mansioni necessarie per selezionare, migliorare, rendere i dati interpretabili.

Il lavoro digitale di questi proletari del clic è fondamentale per produrre quella che spesso non è altro che intelligenza artificiale "fatta a mano".

Il mercato del microlavoro coinvolge oggigiorno un numero crescente di persone. Si stima la sua estensione tra un minimo di quaranta e un massimo di alcune centinaia di milioni di lavoratori.

Antonio A. Casilli, *Schiavi del clic. Perché lavoriamo tutti per il nuovo capitalismo?*, Milano, Feltrinelli, 2020.

# Colonialismo digitale

BUSINESS • TECHNOLOGY

Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic

https://time.com/6247678/openai-chatgpt-kenya-workers/

*ChatGpt, Antonio Casilli: Il lato oscuro dell'algoritmo è la forza lavoro*, «Il Manifesto», 1 aprile 2023, https://ilmanifesto.it/chatgpt-il-lato-oscuro-dellalgoritmo-e-la-forza-lavoro-che-addestra-la-macchina-digitale

# Intelligenza artificiale e discriminazioni

Le decisioni basate sui sistemi di apprendimento automatico sono costitutivamente discriminatorie, in quanto procedono etichettando le persone e raggruppandole in varie classi, secondo qualsiasi tipo di regolarità nei dati di partenza.

Essendo radicata nella natura statistica di questi sistemi, la caratteristica di dimenticare i "margini" è strutturale: non è accidentale e non è dovuta a *bias* singoli e tecnicamente modificabili.

Le decisioni algoritmiche

- replicano attraverso l'automazione le discriminazioni e le disuguaglianze del passato;
- allo stesso tempo, poiché i loro modelli si basano su mere correlazioni, generano nuove e imprevedibili discriminazioni sulla base di fattori irrilevanti.

A. Birhane, E. Ruane, T. Laurent, M.S. Brown, J. Flowers, A. Ventresque, C.L., Dancy, *The Forgotten Margins of AI Ethics*, in *Conference on Fairness, Accountability, and Transparency* (FAccT '22), cit., https://doi.org/10.1145/3531146.3533157.

# Le discriminazioni come diversivo

Discriminazioni contro:

- gruppi tutelati dalla legge;
- "gruppi algoritmici", non previste dalla legge in virtù della loro insensatezza (es. proprietari di cani, adolescenti tristi, video giocatori);
- gruppi creati sulla base di caratteristiche, come la configurazione dei pixel in una foto o il mero ordine di presentazione dei dati, che non sono significativamente riconducibili a individui e sulla base dei quali, invece, possono avvenire trattamenti differenziati.

Parlare di discriminazioni distrae dal chiedersi, di volta in volta, a cosa dovrebbe servire un certo sistema, se funzioni, a chi serva in realtà, che cosa faccia e a chi.

J. Powles, H. Nissenbaum, *The Seductive Diversion of 'Solving' Bias in Artificial Intelligence*, «OneZero», December 7, 2018, https://onezero.medium.com/890df5e5ef53

# Intelligenza artificiale e pensiero magico

Le narrazioni sull'intelligenza artificiale presentano tre caratteristiche proprie del pensiero magico:

1. l'animismo: la tendenza a concepire in termini antropomorfici alcuni oggetti della tecnologia;
2. la mossa, da prestigiatori, di mostrare un risultato, o un effetto, nascondendone allo stesso tempo le cause concrete e i costi;
3. l'assunto che sia possibile prevedere il comportamento futuro di ogni singola persona (come i responsi dell'astrologia, le predizioni algoritmiche a proposito di future azioni o prestazioni umane sono fondate su correlazioni alle quali non corrisponde alcun nesso causale).

D. Tafani, *What's wrong with "AI ethics" narratives*, in «Bollettino telematico di filosofia politica», 2022, pp. 1-22, https://commentbfp.sp.unipi.it/daniela-tafani-what-s-wrong-with-ai-ethics-narratives

# Intelligenza artificiale e cattura culturale

Cattura del regolatore (*regulatory capture*) e cattura culturale (*cultural capture*): colonizzazione dello spazio dell'intermediazione scientifica, come già accaduto con l'industria del tabacco e quella dei combustibili fossili.

I detentori di grandi monopoli intellettuali finanziano la ricerca sui temi dell'intelligenza artificiale, per assicurarsi che siano coerenti con il loro modello di business

- gli esiti
- l'inquadramento teorico
- il tono.

In evidente conflitto di interessi, le grandi aziende tecnologiche finanziano l'ambito di ricerca interdisciplinare dell'etica dell'intelligenza artificiale, che si occupa dei danni e delle sofferenze che esse stesse provocano, con la commercializzazione dei sistemi di IA.

A. Saltelli, D.J. Dankel, M. Di Fiore, N. Holland, M. Pigeon, *Science, the Endless Frontier of Regulatory Capture*, in «Futures», 2022, n. 135

# Sistemi fuori controllo o prodotti fuorilegge?

Pinocchio, alla presenza del giudice, raccontò per filo e per segno l'iniqua frode, di cui era stato vittima […]. Allora il giudice, accennando Pinocchio ai giandarmi, disse loro: — Quel povero diavolo è stato derubato di quattro monete d'oro: pigliatelo dunque e mettetelo subito in prigione. —

Carlo Collodi

D. Tafani, *Sistemi fuori controllo o prodotti fuorilegge? La cosiddetta «intelligenza artificiale» e il risveglio del diritto*, «Bollettino telematico di filosofia politica», 2023, pp. 1-13, https://btfp.sp.unipi.it/it/2023/05/sistemi-fuori-controllo-o-prodotti-fuorilegge

# Il paese di Acchiappacitrulli

Fuga dei giganti della tecnologia dalle loro responsabilità per gli effetti dannosi dei sistemi di apprendimento automatico:

- appelli all'eccezionalità delle nuove tecnologie;
- tesi di un vuoto di responsabilità (*responsibility gap*);
- proposta di una responsabilità distribuita anche tra gli utenti e le vittime
- proposta di attribuire un ruolo decisivo agli esseri umani coinvolti nei processi automatizzati (*human on the loop*), affidando a una persona il compito di intervenire, con prontezza fulminea, nei casi di emergenza, o di rettificare il responso imperscrutabile di un sistema automatico.

L'introduzione di un inverosimile controllo umano svolge la funzione di legittimare l'uso di sistemi pericolosi, attribuendo all'essere umano il ruolo di capro espiatorio.

Quanto simili strategie aziendali siano analoghe alle truffe in senso stretto, è manifesto nei casi in cui le vittime siano colpevoli *by design,* come pare sia stato specificamente previsto nei veicoli Tesla attraverso un meccanismo di disattivazione della funzione Autopilot un secondo prima dell'impatto, così da far risultare che il veicolo, al momento dello schianto, fosse affidato al passeggero umano.

# Ricerca o commercio?

Un ulteriore dispositivo per diffondere impunemente prodotti insicuri e incolpare sistematicamente utenti e clienti è l'offuscamento della linea di confine tra la fase di ricerca e sperimentazione e quella della distribuzione e commercializzazione di prodotti.

OpenAI, ad esempio, rende pubblicamente disponibili i propri generatori di linguaggio naturale, in fase sperimentale, ne distribuisce al tempo stesso versioni a pagamento e fissa termini d'uso in virtù dei quali ogni utente è responsabile, secondo una logica analoga a quella del paese di Acchiappacitrulli, tanto dei propri input quanto degli output prodotti dal sistema (sui quali, ovviamente, non ha alcuna facoltà di controllo o decisione e i cui fondamenti gli sono del tutto oscuri).

# Il principio dell'inevitabilità tecnologica

Il progresso tecnologico è considerato inarrestabile e di ogni singola tecnologia si dà per scontato che sia «qui per restare» o che «se non lo facciamo noi, lo farà qualcun altro». Qualsiasi dibattito ha luogo perciò entro la «logica del fatto compiuto», così che la possibilità di non costruire affatto alcuni sistemi o di non utilizzarli per alcune finalità non possa essere neppure contemplata.

«Il mito dell'inevitabilità tecnologica, politica e sociale è un potente tranquillante per la coscienza. Il suo servizio è quello di togliere la responsabilità dalle spalle di tutti coloro che ci credono veramente. Ma, in realtà, ci sono degli attori! [...] La reificazione di sistemi complessi che non hanno autori, di cui sappiamo solo che ci sono stati dati in qualche modo dalla scienza e che parlano con la sua autorità, non permette di porsi domande di verità o di giustizia.»

Joseph Weizenbaum, *Computer Power and Human Reason. From Judgement to Calculation*, San Francisco, W.H. Freeman & Company, 1976, pp. 241, 252, https://archive.org/details/computerpowerhum0000weiz_v0i3

# Il principio di innovazione

Il principio di innovazione – fondato sul tacito assunto che qualsiasi innovazione tecnologica sia foriera di competitività e occupazione e debba perciò essere assecondata, anche a scapito del principio di precauzione – è in genere la maschera dietro la quale grandi aziende rivendicano la tutela dei loro concreti interessi economici.

Gli studi più recenti mostrano, del resto, che i monopoli del capitalismo intellettuale ostacolano qualsiasi innovazione, per quanto dirompente e benefica, che non si adatti al loro modello di business, promuovendo principalmente un'innovazione tossica che estrae o distrugge valore, anziché produrlo.

A. Saltelli, D.J. Dankel, M. Di Fiore, N. Holland, M. Pigeon, *Science, the Endless Frontier of Regulatory Capture*, in «Futures», 2022, n. 135

# L'animismo

La tendenza umana a concepire gli oggetti della tecnologia in termini antropomorfici è nota e tuttavia irresistibile: risposte emotive e sociali sono generate automaticamente anche da mezzi di comunicazione, quali i televisori o i computer.

Tendiamo a inferire, da una singola, limitata prestazione di un sistema di intelligenza artificiale, che il sistema possieda le competenze che sarebbe ragionevole ascrivere a una persona in grado di svolgere lo stesso compito.

Un generatore di linguaggio naturale, ad esempio, produce stringhe di testo manipolando forme linguistiche – senza accesso al loro significato – sulle base dei modelli statistici costruiti a partire da grandi insiemi di testi digitali: non comprende dunque ciò che scriviamo o ciò che esso stesso scrive più di quanto non lo comprendesse la nostra vecchia macchina da scrivere.

Poiché tuttavia produce testi che appaiono, agli esseri umani, come per lo più dotati di senso compiuto e pertinenti rispetto agli input, o gli si ascrive la capacità di comprendere il linguaggio umano oppure si immagina che la comprensione sia il passo successivo.

# Antropomorfizzazione e deumanizzazione

La tendenza spontanea ad antropomorfizzare gli oggetti della tecnologia è sfruttata dalle aziende per indurre la convinzione che i loro prodotti siano in grado di comprendere, conoscere, consigliare o giudicare in senso proprio.

Correlativamente, il cervello è assimilato a un computer, gli esseri umani a macchine e i volti delle persone a codici a barre, così da giustificare il fatto che le persone siano effettivamente trattate come cose, nei processi in cui si decidono le loro sorti sulla base di statistiche automatizzate, a partire da etichette e collezioni di dati che – a qualsiasi titolo e pur senza alcun riferimento a contesti e significati – li riguardino.

https://twitter.com › sama › status › 1599471830255177728

# I miti dell'eccezionalismo tecnologico e del vuoto giuridico

La tesi che le leggi vigenti non si applichino ai prodotti basati su sistemi di «intelligenza artificiale», in virtù della loro novità e straordinarietà, e che servano dunque nuove leggi, scritte ad hoc per ciascuna tecnologia, serve a dar luogo a una corsa che vedrà il legislatore perennemente in affanno, nel rincorrere le più recenti novità tecnologiche, le quali saranno dunque commercializzabili eslege.

Il problema non è l'assenza di una risposta giuridica, ma piuttosto il dubbio che può esistere sul significato di tale risposta (sarebbe allora più corretto parlare di vaghezza giuridica, in particolare in assenza di precedenti).

D. Tafani, *Sistemi fuori controllo o prodotti fuori legge? La cosiddetta «intelligenza artificiale e il risveglio del diritto*, https://btfp.sp.unipi.it/it/2023/05/sistemi-fuori-controllo-o-prodotti-fuorilegge/

# L’etica dell’intelligenza artificiale come via di fuga dalla regolazione giuridica

La famiglia di narrazioni diffuse dai giganti tecnologici - chiamata "etica dell'IA" - rimuove l'idea che etichettare le persone come cose e trattarle come tali equivalga a negare loro il riconoscimento di qualsiasi diritto, danneggiando infallibilmente gli individui più deboli, e debba quindi essere vietato.

Invece di rifiutare semplicemente le decisioni statistiche automatizzate, le narrazioni sull’etica dell’IA riformulano la questione come un problema di equità algoritmica e di allineamento dei valori, come se

- l'unico problema fosse costituito da pregiudizi singoli ed emendabili;
- gli algoritmi potessero essere dotati delle competenze umane necessarie per formulare giudizi morali;
- i valori morali incorporati nei sistemi di ML potessero essere semplicemente scelti dagli ingegneri e tradotti in termini computazionali.

B. Wagner, *Ethics As An Escape From Regulation. From “Ethics-Washing” To Ethics-Shopping?*, in *Being Profiled: Cogitas Ergo Sum*, ed. by E. Bayamlioglu, I. Baraliuc, L.A.W. Janssens, M. Hildebrandt, Amsterdam, Amsterdam University Press, 2018, pp. 84-89.

# L’istituzionalizzazione aziendale dell’etica

Presa sul serio, l'etica dell'IA richiederebbe un'intelligenza artificiale generale (AGI).

In assenza di AGI, l'equità algoritmica e l'allineamento dei valori non possono essere altro che una strategia di comunicazione aziendale, un *ethics washing*, ossia uno strumento di distrazione per evitare la regolamentazione legale e proseguire il proprio *business as usual*.

Non si tratta dunque di un dibattito teorico, ma di una strategia aziendale.

J. Metcalf, E. Moss, D. Boyd, *Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics*, «Social Research: An International Quarterly», LXXXII, 2, 2019, pp. 449-476, https://datasociety.net/wp-content/uploads/2019/09/Owning-Ethics-PDF-version-2.pdf.

# L'"etica" come merce

Politiche neoliberali di definanziamento pubblico delle università.

Mercato con un'elevata domanda di «prodotti etici», ossia di ricerche sull'etica dalle caratteristiche e dagli esiti prefissati: i ricercatori diventano così «i fornitori di un servizio in questa nuova economia della virtù» e sono indotti alla «complicità con sistemi e attori che cercano di operazionalizzare l'etica per proteggere i propri interessi».

L'«etica dell'intelligenza artificiale» è assimilabile dunque a una merce, a «un'esca per catturare la fiducia» dei cittadini, un mero discorso.

La funzione di tale discorso è quella di tutelare un modello di business fondato sulla sorveglianza e sulla raccolta di dati e metadati individuali.

Non è l'azienda che diventa etica, ma l'«etica» (ossia un mero discorso sull'etica) che diventa un asset aziendale.

*Economies of Virtue: The Circulation of 'Ethics' in AI*, a cura di T. Phan, J. Goldenfein, D. Kuch, M. Mann, Amsterdam, Institute of Network Cultures, 2022, https://networkcultures.org/blog/publication/economies-of-virtue-the-circulation-of-ethics-in-ai/

# La «bolla giuridica»

«Bolla giuridica» (Marco Giraudo): le grandi compagnie tecnologiche hanno fondato il loro modello di business sull'appropriazione e la commercializzazione dei dati personali, in violazione di diritti giuridicamente tutelati, scommettendo su un successivo "salvataggio giuridico", in nome dell'inarrestabilità dell'innovazione tecnologica.

https://www.cambridge.org/core/journals/journal-of-institutional-economics/article/on-legal-bubbles-some-thoughts-on-legal-shockwaves-at-the-core-of-the-digital-economy/62964F5EE993E8A7CC60EB4833FA71CA#article

# «Automaticamente illegali»

**Illegalità dei sistemi di intelligenza artificiale, fino a prova contraria**: è la proposta di Frank Pasquale e Gianclaudio Malgieri, fondata sull'evidenza delle violazioni dei diritti individuali che hanno luogo quando i sistemi di intelligenza artificiale sono utilizzati per ottenere classificazioni o produrre decisioni che hanno effetti rilevanti sulle vite delle persone.

I modelli di IA ad alto rischio incorporati oggi in prodotti e servizi dovrebbero essere disciplinati entro un sistema di "illegalità di default": fino a prova contraria, tali sistemi dovrebbero essere considerati illegali. Prima di immettere sul mercato un prodotto o un servizio che incorpori sistemi di IA ad alto rischio, le aziende avrebbero l'obbligo di dimostrare che la loro tecnologia non è discriminatoria, non è manipolatoria, non è iniqua, non è inaccurata e non è illegittima nelle sue basi giuridiche e nei suoi scopi.

https://brusselsprivacyhub.com/2022/06/01/brussels-privacy-hub-working-paper-from-transparency-to-justification-toward-ex-ante-accountability-for-ai-by-gianclaudio-malgieri-and-frank-pasquale-has-been-published/

# Il risveglio del diritto

In una Dichiarazione congiunta, la Federal Trade Commission e altre tre agenzie federali statunitensi hanno reso noto che, negli interventi di applicazione delle leggi vigenti, esse intendono utilizzare «con vigore» le loro prerogative «per proteggere i diritti degli individui, indipendentemente dal fatto che le violazioni del diritto avvengano attraverso mezzi tradizionali o tecnologie avanzate».

La pretesa che i nuovi sistemi di intelligenza artificiale siano emersi in un «vuoto giuridico» è respinta come infondata, sulla base della elementare constatazione che i codici non prevedono che l'intelligenza artificiale sia esentata dalle leggi.

Quanto al generale non funzionamento di intere categorie di prodotti che le aziende presentano come fondati su sistemi di «intelligenza artificiale», la Commissione esprime la propria consapevolezza di tali pratiche e la decisione di assimilarle a pubblicità ingannevole o a frode.

https://www.ftc.gov/system/files/ftc_gov/pdf/remarks-of-chair-lina-m-khan-re-joint-interagency-statement-on-ai.pdf

“There is no AI exemption to the laws on the books”

Lina M. Khan

Grazie per l'attenzione

daniela.tafani@unipi.it