Bollettino telematico di filosofia politica
Open peer review, put to the test

Daniela Tafani

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

> You must be careful what opinion you hold now-a-days; because if you once allow a particular section of persons to get enough power or get rich, you find them immediately acquiring an extraordinary power of creating your opinions, of forming your mind for you.
>
> Bernard Shaw

Abstract
Giudizi e decisioni che hanno effetti rilevanti sulle vite delle persone 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, utilizzati impropriamente per compiti che non è loro possibile svolgere o che sono impossibili *tout court*.
Le decisioni basate su tali sistemi sono costitutivamente discriminatorie, e dunque, in alcuni ambiti, infallibilmente lesive di diritti giuridicamente tutelati, in quanto procedono trattando gli individui in base al loro raggruppamento in classi, costituite a partire dalle regolarità rilevate 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 singoli *bias* tecnicamente modificabili. Ci si può trovare ai margini dei modelli algoritmici di normalità in virtù di caratteristiche totalmente irrilevanti, rispetto alle decisioni di cui si è oggetto.
Alla vasta documentazione degli esiti ingiusti, nocivi e assurdi di tali decisioni, le grandi aziende tecnologiche – paventando un divieto generalizzato – hanno risposto, in evidente conflitto di interessi, con un discorso sull'etica: è nata così, come operazione di cattura culturale, con l'obiettivo di rendere plausibile un regime di mera autoregolazione, l'«etica dell'intelligenza artificiale». Si tratta di una narrazione che le aziende commissionano e acquistano perché è loro utile come capitale reputazionale, che genera un vantaggio competitivo; è cioè un discorso, rispetto al quale le università hanno il ruolo, e l'autonomia, di un megafono; è «un'esca per catturare la fiducia» dei cittadini: un discorso pubblicitario che, in quanto declamato da altri, non appare neppure come tale.
La funzione di tale discorso è quella di tutelare, legittimandolo, un modello di business - fondato sulla sorveglianza e sulla possibilità di esternalizzare impunemente i costi del lavoro, degli effetti ambientali e dei danni sociali- il cui nucleo consiste nella vendita, alle agenzie pubblicitarie, della promessa di un *microtargeting* fondato sulla profilazione algoritmica.
Negli ultimi anni, l'opera di demistificazione della natura meramente discorsiva e del carattere strumentale dell'«etica dell'intelligenza artificiale», che trasforma l'etica nella questione della conformità procedurale a un «anemico set di strumenti» e standard tecnici, è stata così efficace da indurre molti a liquidare come inutile o dannosa – in quanto disarmata alternativa al diritto o vuota retorica aziendale – l'intera filosofia morale.
Al dissolversi della narrazione sull'etica dell'intelligenza artificiale, compare il convitato di pietra ch'essa aveva lo scopo di tenere alla larga: si sostiene infatti ora, da più parti, l'urgenza che ad intervenire, in modo drastico, sia il diritto. L'adozione di sistemi di apprendimento automatico a fini decisionali, in ambiti rilevanti per la vita delle persone, quali il settore giudiziario, educativo o dell'assistenza sociale, equivale infatti alla decisione, in via amministrativa, di istituire delle «zone pressoché prive di diritti umani».
A chi, in nome dell'inarrestabilità dell'innovazione tecnologica, deplori gli interventi giuridici, giova ricordare che il contrasto non è, in realtà, tra il rispetto dei diritti umani e un generico principio di innovazione, ma tra il rispetto dei diritti umani e il modello di business dei grandi monopoli del capitalismo intellettuale.

Keywords
"AI ethics" as smoke and mirrors · AI hype · AI and magical thinking · AI snake oil · Cultural capture · Ethics washing

Come citare questo articolo
Daniela Tafani, *L'«etica» come specchietto per le allodole. Sistemi di intelligenza artificiale e violazioni dei diritti*, in «Bollettino telematico di filosofia politica», 2023, pp. 1-13, https://commentbfp.sp.unipi.it/letica-come-specchietto-per-le-allodole/.

1. Introduzione

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[1]. In settori quali quello giudiziario, dei servizi finanziari, dell'educazione, dei servizi sociali o del reclutamento del personale, l'uso di sistemi di apprendimento automatico (*machine learning*) nei processi di valutazione e decisione ha dato luogo a esiti ingiusti, nocivi e assurdi[2] – come documenta una letteratura ormai sterminata[3] – con conseguenze che si riverberano a lungo, talora per anni, sulle vite delle vittime[4].

Tali malfunzionamenti non sono occasionali e non sono scongiurabili con interventi tecnici[5]: essi rivelano, anzi, il funzionamento ordinario dei sistemi di apprendimento automatico[6], utilizzati impropriamente per compiti che non è loro possibile svolgere o che sono impossibili *tout court*. Considerato il ruolo cruciale di tali sistemi nel modello di business delle grandi aziende tecnologiche e gli enormi profitti che queste ne ricavano, non sorprende che, paventando un divieto generalizzato, esse mirino a sottrarre tali prodotti all'intervento giuridico: è nata così, con l'obiettivo di rendere plausibile un regime di mera autoregolazione, l'«etica dell'intelligenza artificiale»[7].

Si tratta di una narrazione[8] – ossia di un'idea trasmessa nella forma di storie[9] – finanziata dalle grandi compagnie tecnologiche, in modo complementare alle tradizionali attività di *lobbying*, al fine di scongiurare, o almeno rinviare, la regolazione giuridica[10]. La «cattura del regolatore» attraverso gli opportuni incentivi è così accompagnata dalla cattura culturale: con un'operazione di propaganda, «colonizzando l'intero spazio dell'intermediazione

---

[1] I.D. Raji, I.E. Kumar, A. Horowitz, A.D. Selbst, *The Fallacy of AI Functionality*, in *Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21–24, 2022, Seoul, Republic of Korea*, New York, ACM, 2022, https://doi.org/10.1145/3531146.3533158 (ultimo accesso, a questo e agli altri indirizzi Internet, il 2 aprile 2023).

[2] A. Alkhatib, *To Live in Their Utopia: Why Algorithmic Systems Create Absurd Outcomes*, in *Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan*, New York, ACM, 2021, https://ali-alkhatib.com/papers/chi/utopia/utopia.pdf.

[3] C. O'Neil, *Armi di distruzione matematica. Come i big data aumentano la disuguaglianza e minacciano la democrazia*, trad. it. di D. Cavallini, Milano, Bompiani, 2017; S. U. Noble, *Algorithms of oppression,* New York, New York University Press, 2018; V. Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, New York, St. Martin's Press, 2018. Ne dà atto il governo americano, in White House Office of Science and Technology Policy, *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People*, October 2022, https://www.whitehouse.gov/ostp/ai-bill-of-rights: «Tra le grandi sfide poste oggi alla democrazia c'è l'uso della tecnologia, dei dati e dei sistemi automatizzati in modi che minacciano i diritti dei cittadini americani. Troppo spesso questi strumenti vengono utilizzati per limitare le nostre opportunità e impedire il nostro accesso a risorse o servizi cruciali. Questi problemi sono ben documentati».

[4] V. ad es. A. James, A. Whelan, *'Ethical' artificial intelligence in the welfare state: Discourse and discrepancy in Australian social services,* «Critical Social Policy», XLII, 1, 2022, pp. 22-42, https://journals.sagepub.com/doi/abs/10.1177/0261018320985463.

[5] M. Broussard, *More than a Glitch. Confronting Race, Gender, and Ability Bias in Tech*, Cambridge, Massachusetts, MIT Press, 2023.

[6] Cfr. L. Amoore, *Cloud Ethics. Algorithms and the Attributes of Ourseves and Others*, Durham and London, Duke University Press, 2020, pp. 115-119.

[7] R. Ochigame, *The Invention of "Ethical AI". How Big Tech Manipulates Academia to Avoid Regulation*, «The Intercept», December 20, 2019, https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/.

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

[9] M. D'Eramo, *Dominio. La Guerra invisibile dei potenti contro i sudditi*, Milano, Feltrinelli, 2020.

[10] 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, https://www.degruyter.com/document/doi/10.1515/9789048550180-016/html.

scientifica»[11], si ottiene che il regolatore – al pari dell'opinione pubblica – condivida in partenza l'impostazione desiderata e che chiunque esprima preoccupazioni sia etichettato come retrogrado o luddista[12].
Negli ultimi anni, le voci critiche verso l'etica dell'intelligenza artificiale, dapprima isolate, si sono fatte coro: l'opera di demistificazione del carattere strumentale e meramente discorsivo dell'«etica dell'intelligenza artificiale» è stata così efficace da indurre molti a un «pestaggio dell'etica» (*ethics bashing*), ossia a liquidare come inutile o dannosa – in quanto disarmata alternativa al diritto o vuota retorica aziendale – l'intera filosofia morale[13].
Nel paragrafo che segue, sono esaminate le caratteristiche di una categoria di giudizi e decisioni per i quali non è sensato utilizzare sistemi di apprendimento automatico; in quello successivo, sono presentati e discussi gli elementi ricorrenti – in virtù del conflitto di interessi che le contraddistingue – nelle narrazioni sull'«etica dell'intelligenza artificiale», nonché i più recenti interventi critici. Da ultimo, infine, si rileva la convergenza di studiosi di discipline diverse nella tesi che l'impiego dei sistemi di apprendimento automatico, quando si tratti di assumere decisioni dagli effetti significativi sulle vite delle persone, sia da considerarsi illegittimo, in quanto intrinsecamente irragionevole e infallibilmente lesivo di diritti giuridicamente tutelati.

2. Le proprietà magiche dell'intelligenza artificiale: l'«olio di serpente»

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»[14] – l'apprendimento automatico ha reso possibile, per alcuni specifici compiti, un rapido e genuino progresso. Funzioni quali la previsione e la generazione di stringhe di testo, il riconoscimento facciale, la ricerca per immagini o l'identificazione di contenuti musicali, non trattabili con l'intelligenza artificiale simbolica (giacché non siamo in grado di esplicitare le regole per lo svolgimento di tali compiti e di enumerare esaustivamente i fattori di volta in volta rilevanti), sono affrontate con crescente successo dai sistemi di «apprendimento profondo» (*deep learning*)[15]. Tali sistemi, di natura sostanzialmente statistica, consentono infatti di costruire modelli a partire da esempi, in un processo iterativo di minimizzazione della distanza rispetto ai risultati attesi[16], purché si abbiano a disposizione potenti infrastrutture computazionali e enormi quantità di dati. Le grandi aziende tecnologiche che, intorno al 2010, in virtù di un modello di business fondato sulla sorveglianza[17], 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

---

[11] Sull'«etica dell'intelligenza artificiale» come cattura culturale (*cultural capture*), ossia quale attività che mira a convincere il regolatore, in modo complementare alle tradizionale azioni di *lobbying* o di cattura del regolatore (*regulatory capture*), v. l'ottimo A. Saltelli, D.J. Dankel, M. Di Fiore, N. Holland, M. Pigeon, *Science, the endless frontier of regulatory capture*, «Futures», CXXXV, 2022, pp. 1-14, https://doi.org/10.1016/j.futures.2021.102860.
[12] S. Foucart, S. Horel, S. Laurens, *Les gardiens de la raison. Enquête sur la désinformation scientifique*, Paris, Éditions La Découverte, 2020.
[13] E. Bietti, *From Ethics Washing to Ethics Bashing: A View on Tech Ethics from Within Moral Philosophy*, 2021, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3914119.
[14] M. Gabbrielli, *Dalla logica al deep learning, una breve riflessione sull'intelligenza artificiale*, in *XXVI Lezioni di diritto dell'intelligenza artificiale*, a cura di U. Ruffolo, Torino, Giappichelli, 2021, pp. 3-12: 4.
[15] Y. LeCun, Y. Bengio, G. Hinton, *Deep learning*, DXXI, «Nature», 2015, pp. 436-444.
[16] S. Barocas, M. Hardt, A. Narayanan, *Fairness and Machine Learning: Limitations and Opportunities*, 2019, http://www.fairmlbook.org.
[17] S. Zuboff, *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, trad. it. di P. Bassotti, Roma, Luiss University Press, 2019; C. Doctorow, *How to Destroy Surveillance Capitalism*, «OneZero», August 26, 2020, https://onezero.medium.com/how-to-destroy-surveillance-capitalism-8135e6744d59.

perciò raggiungere, con l'applicazione di algoritmi in gran parte già noti da decenni, traguardi sorprendenti[18].

I rapidi risultati ottenuti in alcune specifiche applicazioni sono dovuti all'adozione di tre scorciatoie: la sostituzione dei nessi causali con mere correlazioni (ossia la rinuncia alla modellazione esplicita e alla possibilità di distinguere le correlazioni spurie dalle relazioni causali[19]), la sostituzione di dati curati puntualmente con dati carpiti o estorti, prelevati in blocco così come si trovano, e, infine, la sostituzione delle variabili rilevanti con *proxy* meno costose, tratte da «campioni di comportamento umano, spesso sotto forma di microscelte effettuate da milioni di utenti»[20]. Tali scorciatoie hanno un costo: simili sistemi sono costitutivamente opachi e fragili, ossia, per gli esseri umani che li utilizzano, sensibili a elementi irrilevanti e soggetti a esiti imprevedibili e assurdi[21].

Le grandi aziende tecnologiche hanno colto comunque l'opportunità di un'espansione illimitata di prodotti e servizi «intelligenti», sfruttando la tendenza umana a concepire in termini antropomorfici gli oggetti della tecnologia e l'effettiva novità di sistemi in grado di svolgere ora, con la forza bruta delle statistiche automatizzate, alcuni compiti prima solo umani: 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[22]? 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 il futuro rendimento scolastico di ogni singolo studente[23]?

Con la promessa di prestazioni di tal genere, fondata sul «dogma che tutti i problemi siano uguali»[24] – nonché sul recupero di antiche pseudoscienze, quali la fisiognomica o la frenologia[25], e sull'invenzione di nuove, quali la psicografia[26] – sistemi di «intelligenza artificiale» sono oggi venduti a enti pubblici e privati, che li utilizzano per comminare pene, allocare risorse, concedere o negare opportunità rilevanti alle persone. La pratica di amplificare oltre misura le presunte prestazioni dei sistemi di intelligenza artificiale (*AI hype*)[27], traendo

---

[18] *'Open Secrets': An Interview with Meredith Whittaker*, in *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/, pp. 140-152: 145.

[19] C.S. Calude, G. Longo, *The Deluge of Spurious Correlations in Big Data*, «Foundations of Science», XXII, 2017, pp. 595–612, https://www.di.ens.fr/users/longo/files/BigData-Calude-LongoAug21.pdf.

[20] N. Cristianini, *Shortcuts to Artificial Intelligence*, in *Machines We Trust. Perspectives on Dependable AI,* ed. by M. Pelillo, T. Scantamburlo, Cambridge, Massachusetts, The MIT Press, 2021, pp. 11-25, https://philpapers.org/archive/CRISTA-3.pdf; Idem, *La scorciatoia. Come le macchine sono diventate intelligenti senza pensare in modo umano*, Bologna, Il Mulino, 2023.

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

[22] I.K. Williams, *Can A.I.-Driven Voice Analysis Help Identify Mental Disorders?*, «The New York Times», April 5, 2022, https://www.nytimes.com/2022/04/05/technology/ai-voice-analysis-mental-health.html.

[23] Recepisce acriticamente tale narrazione un recente documento della Commissione Europea (*Orientamenti etici per gli educatori sull'uso dell'intelligenza artificiale (IA) e dei dati nell'insegnamento e nell'apprendimento,* Lussemburgo, Ufficio delle pubblicazioni dell'Unione europea, 2022, https://op.europa.eu/it/publication-detail/-/publication/d81a0d54-5348-11ed-92ed-01aa75ed71a1/language-it).

[24] A. Narayanan, S. Kapoor, *Why are deep learning technologists so overconfident?*, August 31, 2022, https://aisnakeoil.substack.com/p/why-are-deep-learning-technologists.

[25] K. Crawford, *Né intelligente né artificiale. Il lato oscuro dell'IA*, Bologna, Il Mulino, 2021; L. Stark, J. Hutson, *Physiognomic Artificial Intelligence*, «Fordham Intellectual Property, Media and Entertainment Law Journal», XXXII, 4, 2022, pp. 922-978, https://ir.lawnet.fordham.edu/iplj/vol32/iss4/2.

[26] A.G. Martínez, *The Noisy Fallacies of Psychographic Targeting*, «Wired», March 19, 2018, https://www.wired.com/story/the-noisy-fallacies-of-psychographic-targeting/.

[27] V., da ultimo, I. Van Rooij, *Stop feeding the hype and start resisting*, 2022, https://irisvanrooijcogsci.com/2023/01/14/stop-feeding-the-hype-and-start-resisting/.

esempi, con una fallacia logica, dal futuro o dalla fantascienza[28], costituisce un atto di persuasione[29], per lo più a fini di marketing, e un esercizio di potere[30], giacché la decisione di un sistema di apprendimento automatico non consente spiegazioni e non ammette ricorsi. L'abuso della credulità popolare è ormai così plateale da aver dato luogo all'espressione «intelligenza artificiale olio di serpente» (*AI snake oil*), in memoria di quell'intruglio a base di trementina, olio minerale, grasso di cottura, peperoncino, diluente per vernici e insetticida, che il cowboy Clark Stanley vendeva ai gonzi nel Far West (con la raccomandazione di diffidare delle imitazioni) come taumaturgico rimedio per tutti i mali[31].

Negli Stati Uniti, la Federal Trade Commission ha denunciato «il problema della falsa intelligenza artificiale»[32] – ossia la diffusa pratica aziendale di utilizzare l'espressione «intelligenza artificiale», che è oggi «un termine di marketing», come «argomento ingannevole per vendere nuovi prodotti e servizi» – e ha diffidato le aziende dal sostenere che i loro prodotti di IA possano «fare qualcosa che va oltre le attuali capacità di qualsiasi IA o tecnologia automatizzata»: «ad esempio, non viviamo ancora nel regno della fantascienza, in cui i computer siano in grado fare [...] previsioni affidabili del comportamento umano». Alle aziende che avanzino pretese infondate, quanto alle prestazioni dei loro prodotti di IA, non serve una macchina, ricorda la Federal Trade Commission, per predire cosa la medesima Commissione potrebbe fare[33].

Tra i sistemi di apprendimento automatico che hanno la stessa natura, e un pari successo, dell'«olio di serpente», destano un interessato entusiasmo i sistemi di «ottimizzazione predittiva», ossia gli algoritmi di decisione fondati su previsioni circa il futuro di singoli individui. Sono in commercio e in uso, ormai da anni, sistemi che consentono di assumere decisioni in modo automatico, grazie alla possibilità di prevedere – si dice – se un cittadino commetterà un crimine[34], se un candidato per un impiego sarà efficiente e collaborativo, se uno studente abbandonerà gli studi, se un minore sarà maltrattato dai suoi familiari, se una determinata persona restituirà il prestito eventualmente concessole o se avrà bisogno di specifica assistenza medica[35].

La tesi che i sistemi di apprendimento automatico siano in grado di prevedere eventi o azioni future di singole persone non ha alcun fondamento scientifico. La convinzione che ciò sia possibile si fonda, come l'astrologia, su una commistione di matematica e superstizione, e, in particolare, sull'idea – caratteristica della superstizione e ascritta, nel XX secolo, al mondo della

---

[28] G. Musa, *Echoes of myth and magic in the language of Artificial Intelligence*, «AI & SOCIETY», XXXV, 4, 2020, https://link.springer.com/article/10.1007/s00146-020-00966-4; S.-A. Hong, *Predictions without futures*, «Historical Futures», 2022, https://onlinelibrary.wiley.com/doi/10.1111/hith.12269.

[29] J. Stilgoe, *Who's Driving Innovation? New Technologies and the Collaborative State*, Cham, Palgrave Macmillan, 2020, pp. 40-41.

[30] P.R. Lewis, S. Marsh, J. Pitt, *AI vs «AI»: Synthetic Minds or Speech Acts,* «IEEE Technology and Society Magazine», 2021, pp. 6-13, https://ieeexplore.ieee.org/document/9445758.

[31] C. Stanley, *The Life and Adventures of the American Cow-Boy, 1897,* https://archive.org/details/F596S822CowboyImages.

[32] M. *Atleson, Chatbots, deepfakes, and voice clones: AI deception for sale*, March 20, 2023, https://www.ftc.gov/business-guidance/blog/2023/03/chatbots-deepfakes-voice-clones-ai-deception-sale.

[33] M. Atleson, *Keep your AI claims in check*, Febraury 27, 2023, https://www.ftc.gov/business-guidance/blog/2023/02/keep-your-ai-claims-check.

[34] Per un recente utilizzo di un sistema di polizia predittiva, in Germania, v. M. Monroy, *Klage gegen Kommissar Computer*, «nd», 20. Dezember 2022, https://www.nd-aktuell.de/artikel/1169495.datenauswertung-bei-der-polizei-klage-gegen-kommissar-computer.html. Sulle violazioni di diritti giuridicamente tutelati che la polizia predittiva costitutivamente comporta, agendo sulla base di una «pericolosità» individuale stimata su basi meramente statistiche, a partire dalla totalità dei dati che riguardino, a qualsiasi titolo, un cittadino, v. Frank Pasquale, *Le nuove leggi della robotica. Difendere la competenza umana nell'era dell'intelligenza artificiale*, trad. it. di P. Bassotti, Roma, LUISS University Press, 2021, p. 152.

[35] 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.

psicosi[36] – che tutte le connessioni siano significative, indipendentemente dalla presenza di nessi causali, e che tutto spieghi tutto[37].
Come la fede nei pronostici dell'astrologia, la credenza nelle previsioni algoritmiche svanisce non appena si applichino i criteri di comunicabilità e riproducibilità propri della scienza moderna[38]. 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, quando si debbano allocare risorse scarse e non sia possibile utilizzare – anziché sistemi di apprendimento automatico – semplici procedimenti di calcolo con variabili pertinenti ed esplicite[39].
I sistemi di apprendimento automatico procedono raggruppando i singoli individui in classi, entro un processo, costitutivamente opaco, di discriminazione in senso tecnico. Le decisioni fondate su tali sistemi sono strutturalmente discriminatorie: esse assumono infatti che «tutto quello che è accaduto in passato si ripeterà» e che «quello che fanno persone categorizzate come simili ad altre tenderà ad estendersi a tutti gli altri membri dello stesso cluster»[40].
Se la previsione è utilizzata per mostrare annunci pubblicitari, la cosiddetta previsione «personalizzata» è «una sorta di profezia digitale che è solo leggermente più accurata rispetto a metodi analogici come la lettura della mano» e costituisce in realtà, a partire dall'omologazione del singolo alla classe alla quale lo si è ascritto, un meccanismo di manipolazione:

> Quando si va a scavare tra i reali meccanismi tecnici che calcolano la prevedibilità, si arriva a capire che i suoi fondamenti scientifici sono di fatto antiscientifici, e che lo stesso nome con cui definiamo la prevedibilità stessa è fatalmente sbagliato: si tratta soltanto di manipolazione. Se un sito web inizia a dirvi che, poiché vi è piaciuto questo libro, potrebbero piacervi anche i libri di James Clapper o di Michael Hayden non sta offrendo un suggerimento sensato, quanto piuttosto un meccanismo di sottile coercizione.[41]

Quando il responso dei sistemi di apprendimento automatico sia utilizzato a fini decisionali, in ambiti rilevanti per la vita delle persone, quali il settore giudiziario, educativo o dell'assistenza sociale, la decisione produce ciò che pretende di prevedere: se «il genere predice una paga più bassa e il colore della pelle predice la probabilità di essere fermati dalla polizia», con il passaggio dalla previsione alla decisione tale profilazione sociale si autorinforza, automatizzando, con la legittimazione fornita dalla presunta oggettività algoritmica, i pregiudizi incorporati nella descrizione statistica iniziale[42]. Come ha osservato Meredith Broussard, «l'uso delle macchine per la polizia predittiva è efficace quanto spargere sale su una ferita»[43].
Oltre alle discriminazioni contro gruppi specificamente tutelati dal diritto, il passaggio dalla previsione alla prescrizione, senza che siano trasparenti le variabili il cui peso incide sulla previsione, rende automaticamente rilevanti fattori non pertinenti in una decisione sensata: la probabilità di recidiva di un giovane, ad esempio, è ovviamente più alta di quella di un anziano, in virtù del mero numero di anni che presumibilmente gli restano da vivere, ma questo fattore – che un sistema di apprendimento automatico traccerà in modo inestricabile dagli altri – non costituisce, per gli esseri umani, una ragione valida per trattenerlo in carcere; ulteriori fattori

---

[36] P. Rossi, *Il tempo dei maghi. Rinascimento e modernità*, Milano, Raffaello Cortina, 2006, pp. 305-306.
[37] D. Tafani, *What's wrong with «AI ethics» narratives*, cit.
[38] P. Rossi, *la nascita della scienza moderna in Europa,* Roma-Bari, Laterza, 2000, p. xiii.
[39] A. Wang, S. Kapoor, S. Barocas, A. Narayanan, *Against Predictive Optimization: On the Legitimacy of Decision-Making Algorithms that Optimize Predictive Accuracy*, cit.
[40] T. Numerico, *Big data e algoritmi. Prospettive critiche*, Roma, Carocci, 2021, p. 170.
[41] E. Snowden, *Errore di sistema*, Milano, Longanesi, 2019.
[42] D. McQuillan, *Resisting AI. An Anti-fascist Approach to Artificial Intelligence*, Bristol, Bristol University Press, 2022; S.-A. Hong, *Predictions without futures*, cit.
[43] M. Broussard, *More than a Glitch. Confronting Race, Gender, and Ability Bias in Tech*, cit.

rilevanti per la previsione, e tuttavia non ascrivibili alla responsabilità di un imputato, quali il suo luogo di nascita, non possono essere sensatamente annoverati tra le motivazioni di una condanna e saranno invece utilizzati, in modo opaco, da un sistema di apprendimento automatico[44].
La decisione di utilizzare sistemi di apprendimento automatico in ambiti rilevanti per le vite delle persone equivale dunque, di fatto, a una presa di posizione politica a favore dello *status quo*, ossia alla decisione di replicare il passato, automatizzando – ossia, al tempo stesso, mascherando e amplificando - le disuguaglianze e le discriminazioni[45].
Il compito di rimuovere tali discriminazioni algoritmiche, come se la giustizia coincidesse con una replica automatizzata del passato, al netto delle discriminazioni, è stato affidato all'etica dell'intelligenza artificiale.

3. «Etica dell'intelligenza artificiale» e cattura culturale

L'etica dell'intelligenza artificiale è un ambito di ricerca interdisciplinare promosso e finanziato in massima parte dalle grandi aziende tecnologiche, in risposta alle denunce delle discriminazioni, dei danni e delle ingiustizie generati da sistemi prodotti, gestiti e venduti dalle medesime aziende. Il conflitto di interessi[46], come è stato osservato, è analogo a quello che avrebbe luogo se la ricerca sulla mitigazione degli effetti del fumo fosse finanziata e diretta dalle multinazionali del tabacco[47].
Dal 2014 – rileva uno dei più recenti rapporti sull'intelligenza artificiale – la ricerca sull'etica dell'intelligenza artificiale «è esplosa» (con una crescente incidenza dei ricercatori che hanno anche un'affiliazione industriale), la potenza e le prestazioni dei sistemi sono aumentate e, contestualmente, è aumentata, anziché diminuire, la tossicità di tali sistemi, ossia la loro tendenza a riprodurre e amplificare le discriminazioni riflesse nei dati di partenza[48]. Analogamente, le centinaia di linee guida sui principi etici per lo sviluppo e l'applicazione dell'intelligenza artificiale – promosse o approvate da istituzioni, stati nazionali, enti sovranazionali e grandi imprese – si sono rivelate mere dichiarazioni di principio, reticenti su aspetti moralmente cruciali (quali lo sfruttamento del lavoro o gli effetti sull'ambiente) e prive di effetti concreti[49]. Quanto, infine, ai comitati etici aziendali o ai gruppi di ricerca interni istituiti dai grandi oligopolisti in seguito a scandali dalla vasta risonanza e in risposta alla crescente avversione pubblica nei loro confronti (*techlash*), è ormai considerato paradigmatico il caso di Timnit Gebru e Margaret Mitchell, *co-leader* dell'*Ethical AI team* di Google, licenziate da Google per aver rifiutato di ritirare la loro firma da un articolo sui costi ambientali e i danni sociali dei grandi modelli di elaborazione del linguaggio naturale e per non aver accettato

---

[44] A. Wang, S. Kapoor, S. Barocas, A. Narayanan, *Against Predictive Optimization: On the Legitimacy of Decision-Making Algorithms that Optimize Predictive Accuracy*, cit.
[45] T. Numerico, *Big data e algoritmi. Prospettive critiche*, cit., pp. 173-177. D. McQuillan, *Resisting AI. An Anti-fascist Approach to Artificial Intelligence*, cit., p. 43.
[46] O. Williams, *How Big Tech funds the debate on AI ethics*, «The New Statesman», June 6, 2019 (updated June 7, 2021), https://www.newstatesman.com/science-tech/2019/06/how-big-tech-funds-debate-ai-ethics.
[47] M. Abdalla, M. Abdalla, *The Grey Hoodie Project: Big Tobacco, Big Tech, and the threat on academic integrity*, in *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES '21), May 19–21, 2021, Virtual Event*, New York, ACM, 2021, pp. 287-297, https://arxiv.org/abs/2009.13676v4.
[48] D. Zhang *et alii, The AI Index 2022 Annual Report*, AI Index Steering Committee, Stanford Institute for Human-Centered AI, Stanford University, 2022, p. 105, https://aiindex.stanford.edu/report/.
[49] T. Hagendorff, *The Ethics of AI Ethics - An Evaluation of Guidelines*, *Minds and Machines*, XXX, 2020, pp. 99–120, https://link.springer.com/content/pdf/10.1007/s11023-020-09517-8.pdf; L. Munn, *The uselessness of AI ethics*, «AI and Ethics», 2022, https://link.springer.com/article/10.1007/s43681-022-00209-w.

neppure l'invito, da parte di Google, a mitigarne le conclusioni con qualche accenno a imminenti soluzioni tecniche[50].

Da qualche anno, l'«etica dell'intelligenza artificiale» è divenuta perciò, oltre che un ambito disciplinare, essa stessa oggetto di studi critici: di fronte all'evidenza di un'operazione di cattura culturale, volta a sfuggire alla regolazione giuridica per il tramite di un discorso sull'etica e di una promessa di autoregolazione, perfino nei grandi convegni annuali dedicati all'etica dell'intelligenza artificiale sono comparsi interventi sul conflitto di interessi, denunciando la mancanza di credibilità delle ricerche presentate in quegli stessi convegni[51]. Le grandi aziende tecnologiche risultano in effetti finanziatrici dell'intero ambito, dai rinfreschi dei convegni fino ai fondi per la ricerca accademica e ai laboratori, e dirigono le ricerche determinandone l'impostazione, i risultati e perfino il tono[52], affinché siano coerenti con il loro modello di business. Non si tratta dunque di un dibattito teorico, ma di una strategia aziendale, di una «lotta per il potere», in cui le critiche sono riassorbite e neutralizzate, cooptando i critici più deboli e punendo il dissenso[53]. Le politiche neoliberali di definanziamento pubblico delle università e i paralleli incentivi ad ottenere finanziamenti industriali[54] indirizzano le università, e in particolare quelle le cui ricerche richiedono costose infrastrutture computazionali, verso un 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»[55], trasformando l'etica nella questione della conformità procedurale a un «anemico set di strumenti»[56] e standard tecnici.

L'«etica dell'intelligenza artificiale» è assimilabile dunque a una merce, che i ricercatori e le università sono interessati a fornire, in quanto «olio che unge le ruote della collaborazione»[57] con le grandi aziende tecnologiche, e che le aziende commissionano e acquistano perché è loro

---

[50] E.M. Bender, T. Gebru, A. Mc Millan-Major, S. Shmitchell, *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*, in *Conference on Fairness, Accountability, and Transparency (FAccT '21), March 3–10, 2021, Virtual Event, Canada*, New York, ACM, 2021, https://dl.acm.org/doi/10.1145/3442188.3445922. Per una selezione degli articoli sulla vicenda, a cura di Emily Bender, https://faculty.washington.edu/ebender/stochasticparrots/.

[51] M. Young, M.A. Katell, P. M. Krafft, *Confronting Power and Corporate Capture at the FAccT Conference*, in *Conference on Fairness, Accountability, and Transparency (FAccT '22)*, cit., https://dl.acm.org/doi/10.1145/3531146.3533194; B.L. Gansky, S. M. Mcdonald, *CounterFAccTual: How FAccT Undermines Its Organizing Principles*, in *Conference on Fairness, Accountability, and Transparency (FAccT '22)*, cit., https://doi.org/10.1145/3531146.3533241.

[52] P. Dave, J. Dastin, *Google told its scientists to "strike a positive tone" in AI research-documents*, «Reuters», December 23, 2020, https://www.reuters.com/article/us-alphabet-google-research-focus-idUSKBN28X1CB.

[53] M. Whittaker, *The steep cost of capture,* «Interactions», XXVIII, 6, 2021, pp. 51-55, https://interactions.acm.org/archive/view/november-december-2021/the-steep-cost-of-capture.

[54] Per un quadro complessivo, che include gli effetti dei metodi di valutazione della ricerca, v. M.C. Pievatolo, *Integrità della ricerca: i numeri, gli uomini e la scienza*, «Bollettino telematico di filosofia politica», 2018, https://btfp.sp.unipi.it/it/2018/05/uominienumeri/; Idem, *Sulle spalle dei mercanti? Teledidattica e civiltà tecnologica*, 2022, https://zenodo.org/record/6544650.

[55] T. Phan, J. Goldenfein, M. Mann, D. Kuch, *Economies of Virtue: The Circulation of 'Ethics' in Big Tech*, «Science as Culture» (in corso di pubblicazione), https://ssrn.com/abstract=3956318. Per un esempio, invece, di difesa della tesi dell'«etica come servizio» (*Ethics as a Service*) e dell'«operazionalizzazione dell'etica», v. J. Morley, A. Elhalal, F. Garcia, L. Kinsey, J. Mökander, L. Floridi, *Ethics as a Service: A Pragmatic Operationalisation of AI Ethics*, «Minds and Machines», XXXI, 2021, pp. 239–256, https://link.springer.com/article/10.1007/s11023-021-09563-w.

[56] 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.

[57] M. Richardson, *Military Virtues and the Limits of 'Ethics' in AI Research*, in *Economies of Virtue: The Circulation of 'Ethics' in Big Tech,* cit., pp. 130-137: 134.

utile come capitale reputazionale[58], che genera un vantaggio competitivo[59]; è cioè un discorso[60], rispetto al quale le università hanno il ruolo, e l'autonomia, di un megafono; è «un'esca per catturare la fiducia»[61] dei cittadini: un discorso pubblicitario che, in quanto declamato da altri, non appare neppure come tale.
La funzione di tale discorso è quella di tutelare, legittimandolo, un modello di business – fondato sulla sorveglianza e sulla possibilità di esternalizzare impunemente i costi del lavoro, degli effetti ambientali e dei danni sociali – il cui nucleo consiste nella vendita, alle agenzie pubblicitarie, della promessa di un *microtargeting* fondato sulla profilazione algoritmica[62]. Non è dunque l'azienda che diventa etica, ma l'«etica» (ossia un mero discorso sull'etica) che diventa un *asset* aziendale strategico, che consente di «nascondere ingiustizie sistemiche dietro la bandiera della virtù»[63].
Poiché l'impostazione del discorso è dettata dalla sua funzione, l'etica dell'intelligenza artificiale è inquadrata nella prospettiva del determinismo tecnologico e del soluzionismo[64], entro la «logica del fatto compiuto»[65]. Si assume, con una prospettiva soluzionista, che risolvere un problema sociale equivalga a trattarlo con un sistema di apprendimento automatico, trovando un obiettivo rispetto al quale minimizzare la funzione di perdita[66]. La possibilità di non costruire affatto alcuni sistemi o di non utilizzarli per alcune finalità non è mai contemplata, poiché il discorso sull'etica svuota preventivamente di significato le domande sull'opportunità e la legittimità dello sviluppo e dell'applicazione di alcune tecnologie, ponendole «come inevitabili e, purché siano adottati *frameworks* etici, lodevoli»[67]. Le decisioni di business restano così al riparo dalla discussione, tramite un set di problemi e soluzioni che individua nel *design* tecnico il livello appropriato per la soluzione di tutti i problemi[68]. Quando non siano ascrivibili a un emendabile errore di design, le ingiustizie sono ricondotte alle sole malefatte di singoli soggetti (*bad actors*)[69] anziché a rapporti di

[58] J. Metcalf, E. Moss, D. Boyd, *Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics*, cit.
[59] Naturalmente, come è stato osservato, il valore dell'etica come merce è condizionato al suo avere effetti etici assai limitati, che non intralcino il proseguimento del *business as usual*, e alla sua capacità di proteggere le aziende dalle critiche strutturali (T. Phan, J. Goldenfein, M. Mann, D. Kuch, *Introduction: Economies of Virtue*, in *Economies of Virtue: The Circulation of 'Ethics' in AI,* cit., p. 14).
[60] T. Linnet, L. Dencik, *Constructing Commercial Data Ethics*, «Technology and Regulation», 2, 2020, pp. 1-10, https://doi.org/10.26116/techreg.2020.001.
[61] S. Pinker, *Extractivist Ethics*, in *Economies of Virtue: The Circulation of 'Ethics' in Big Tech,* cit., pp. 39-48: 43. V. anche J.J. Bryson, *AI & Global Governance: No One Should Trust AI*, United Nations University, Centre for Policy Research, November 13, 2018, https://cpr.unu.edu/publications/articles/ai-global-governance-no-one-should-trust-ai.html.
[62] C. Doctorow, *How to Destroy Surveillance Capitalism*, cit.
[63] T. Phan, J. Goldenfein, M. Mann, D. Kuch, *Economies of Virtue: The Circulation of 'Ethics' in Big Tech*, cit.
[64] E. Mozorov, *To Save Everything, Click Here: The Folly of Technological Solutionism*, New York, Public Affairs, 2013.
[65] C. Tessier, *Éthique et IA: analyse et discussion*, in *CNIA 2021: Conférence Nationale en Intelligence Artificielle*, a cura di O. Boissier, 2021, pp. 22-29, https://hal-emse.ccsd.cnrs.fr/emse-03278442.
[66] D. McQuillan, *Resisting AI. An Anti-fascist Approach to Artificial Intelligence*, cit., p. 15.
[67] A. James, A. Whelan, *'Ethical' artificial intelligence in the welfare state: Discourse and discrepancy in Australian social services*, cit., p. 37.
[68] Sostiene, tra gli altri, la centralità del *design* L. Floridi, *Etica dell'intelligenza artificiale. Sviluppi, opportunità e sfide*, trad. it. a cura di M. Durante, Milano, Raffaello Cortina, 2022, per il quale la stessa filosofia sarebbe design concettuale. Per un'analisi critica di quest'ultima tesi, v. R. Mordacci, *Sapere, progetto, azione. Sul primato del pratico*, «Phenomenology and Mind», XX, 2021, pp. 176-182, https://doi.org/10.17454/pam-2015.
[69] P. O'Shea, L. Conklin, E. L'Hôte, M. Smirnova, *Communicating About the Social Implications of AI: A FrameWorks Strategic Brief. Artificial intelligence (AI): good or evil?*, 2021, https://www.frameworksinstitute.org/publication/communicating-about-the-social-implications-of-ai-a-frameworks-strategic-brief/

sfruttamento o di potere strutturali[70] o alla decisione politica di trattare le questioni sociali quali problemi di disciplinamento e controllo, passibili di soluzioni tecniche[71].
Il discorso sull'«etica dell'intelligenza artificiale» tematizza infatti la sola, ristretta questione dell'equità algoritmica (assimilando l'opera di moralizzazione degli algoritmi a un obiettivo tecnico plausibile), tacendo, come moralmente non rilevanti, le questioni relative ai danni ambientali[72], al colonialismo dei dati[73] e, in particolare, allo sfruttamento del lavoro[74], sui quali si regge lo sviluppo e l'utilizzo dei sistemi di apprendimento automatico[75]. A quella stessa tecnologia, concepita come immateriale e apolitica, è conferito anzi il potere di risolvere i problemi ch'essa contribuisce pesantemente a creare[76].
All'«etica dell'intelligenza artificiale» è attribuito anche il compito di agevolare i giganti della tecnologia nella fuga dalle loro responsabilità per gli effetti dannosi dei sistemi di apprendimento automatico: le proposte utili a questo scopo – quali quelle di riconoscere, in virtù dell'eccezionalità delle nuove tecnologie, un vuoto di responsabilità (*responsibility gap*)[77] o una responsabilità senza colpa (*faultless responsibility*), distribuita anche tra gli utenti e le vittime[78] – sono state ormai demistificate, in virtù della manifesta implausibilità di un superamento, sulle sole ali della fantascienza, della dicotomia tra soggetti giuridici e oggetti[79].
Infondata e irrealistica è stata giudicata anche la proposta ulteriore, 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, non si sa su quali basi, il responso imperscrutabile di un sistema automatico: l'introduzione di un inverosimile controllo umano svolge – come è stato osservato – la

[70] 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.
[71] S.-A. Hong, *Predictions without futures*, cit.
[72] K. Crawford, *Né intelligente né artificiale. Il lato oscuro dell'IA*, cit..
[73] N. Couldry, U.A. Mejias, *Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject*, «Television & New Media», XX, 4, 2018, pp. 336–49.
[74] A. Mateescu, M.C. Elish, *AI in context, the labor of integrating new technologies*, New York, Data & Society Research Institute, 2019, https://datasociety.net/library/ai-in-context/; A. Aloiso, G. De Stefano, *Il tuo capo è un algoritmo. Contro il lavoro disumano*, Bari-Roma, Laterza, 2020; A.A. Casilli, *Schiavi del clic. Perché lavoriamo tutti per il nuovo capitalismo?*, Milano, Feltrinelli, 2020; S. Mezzadra, *Oltre il riconoscimento. Piattaforme digitali e metamorfosi del lavoro*, «Filosofia politica», XXXV, 3, 2021, pp. 487-502.
[75] V., da ultimo, B. Perrigo, *OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic*, «Time», January 18, 2023, https://time.com/6247678: l'etichettatura dei contenuti «tossici» – che provoca danni alla salute mentale dei lavoratori il cui unico compito consista nel vedere immagini o leggere testi su stupri, torture, pedopornografia e violenze di ogni altro genere, al fine di classificarli – si rende «necessaria» solo perché le aziende adottano la scorciatoia di utilizzare in blocco tutti i dati disponibili o di cui riescono ad appropriarsi, anziché sostenere il costo della generazione, dell'annotazione e della cura di dataset appropriati, privi in partenza di simili contenuti. Su questo aspetto essenziale, della «catena di approvvigionamento dei dati», v. N. Cristianini, *Shortcuts to Artificial Intelligence*, cit.; Idem, *La scorciatoia. Come le macchine sono diventate intelligenti senza pensare in modo umano*, cit.
[76] S. Taffel, L. Bedford, M. Mann, *Ecocide Isn't Ethical: Political Ecology and Capitalist AI Ethics*, in *Economies of Virtue: The Circulation of 'Ethics' in AI,* cit., pp. 58-82.
[77] A. Matthias, *The responsibility gap: Ascribing responsibility for the actions of learning automata,* «Ethics and Information Technology», 2004, n. 6,3, pp. 175–183.
[78] Così L. Floridi, *Faultless responsibility: on the nature and allocation of moral responsibility for distributed moral actions*, «Philosophical Transactions of the Royal Society A», CCCLXXIV, 2083, 2016, https://doi.org/10.1098/rsta.2016.0112.
[79] A. Bertolini, *Artificial intelligence does not exist! Defying the technology- neutrality narrative in the regulation of civil liability for advanced technologies*, «Europa e diritto privato», 2022, n. 2, pp. 369-420, https://www.academia.edu/86737226; v. anche K. Yeung, *A Study of the Implications of Advanced Digital Technologies (Including AI Systems) for the Concept of Responsibility Within a Human Rights Framework*, Council of Europe, 2019, pp. 39-40, 64-65, https://rm.coe.int/a-study-of-the-implications-of-advanced-digital-technologies-including/168094ad40.

funzione di legittimare l'uso di sistemi che restano in realtà fuori controllo[80]. Il ruolo dell'essere umano non può essere dunque che quello di capro espiatorio, come pare sia stato specificamente previsto nei veicoli Tesla (attraverso un meccanismo di disattivazione del pilota automatico meno di un secondo prima dell'impatto)[81].
Quanto alle scelte di vita o di morte che i veicoli a guida autonoma dovrebbero assumere nei casi di incidente inevitabile, in una riedizione di grande successo mediatico del dilemma del *trolley*[82], il recente caso giudiziario che coinvolge Tesla – per aver annunciato, come già in commercio, «veicoli a guida autonoma» che non sono affatto possibili, nel presente stadio di sviluppo della tecnologia – esibisce concretamente l'affinità tra le narrazioni fondate sulla fantascienza e le truffe in senso stretto[83].
Al dissolversi della narrazione sull'etica dell'intelligenza artificiale, compare il convitato di pietra ch'essa aveva lo scopo di tenere alla larga: si sostiene infatti ora, da più parti, l'urgenza che ad intervenire, e in modo drastico, sia il diritto[84].

4. Tutela dei diritti e illegalità di *default* dei sistemi di intelligenza artificiale

Nei primi anni del loro impiego, l'evidenza dei danni prodotti dai sistemi di apprendimento automatico è stata affrontata come un problema di discriminazioni, da risolvere con interventi tecnici. Qualora ci si concentri sulle discriminazioni, pur drammaticamente reali, ai danni di gruppi protetti dalla legge, non si coglie, tuttavia, la natura del problema.[85]
Le decisioni basate sui sistemi di apprendimento automatico sono infatti costitutivamente discriminatorie, in quanto procedono trattando gli individui in base al loro raggruppamento in classi, costituite a partire dalle regolarità rilevate nei dati di partenza. Essendo radicata nella natura statistica di questi sistemi, la caratteristica di dimenticare i «margini»[86] è strutturale: non è accidentale e non è dovuta a singoli *bias* tecnicamente modificabili. Ci si può trovare ai margini dei modelli algoritmici di normalità in virtù di caratteristiche totalmente irrilevanti, rispetto alle decisioni di cui si è oggetto[87]: a qualcuno può accadere, ad esempio, di ottenere un prestito a un tasso di interesse più elevato perché acquista la stessa marca di birra dei debitori insolventi[88] o di essere scartato, in una procedura di selezione del personale, perché sullo

---

[80] B. Green, *The flaws of policies requiring human oversight of government algorithms*, «Computer Law & Security Review», XLV, 2022, https://www.sciencedirect.com/science/article/pii/S0267364922000292.
[81] A. Stoklosa, *NHTSA Finds Teslas Deactivated Autopilot Seconds Before Crashes*, «Motortrend», June 15, 2022, https://www.motortrend.com/news/nhtsa-tesla-autopilot-investigation-shutoff-crash/.
[82] D. Tafani, *Sulla moralità artificiale. Le decisioni delle macchine tra etica e diritto*, «Rivista di filosofia», CXI, 1, 2020, pp. 81-103.
[83] 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.
[84] R. Mcnamee, *Recovering from Big Tech's lost decade*, «Los Angeles Times» January 1, 2023, https://www.latimes.com/opinion/story/2023-01-01/big-tech-went-wrong-pointless-products-and-bad-business-models .
[85] A.L. Hoffmann, *Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse*, «Information, Communication & Society», XXII, 7, 2019, pp. 900-915, https://static1.squarespace.com/static/5b8ab61f697a983fd6b04c38/t/5cd9934e9b747a265111e80a.
[86] 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.
[87] A. Alkhatib, *To Live in Their Utopia: Why Algorithmic Systems Create Absurd Outcomes*, cit.
[88] F. Pasquale, *Le nuove leggi della robotica. Difendere la competenza umana nell'era dell'intelligenza artificiale*, cit., p. 163.

sfondo del video di autopresentazione compare un muro bianco, anziché una libreria o una parete con un quadro[89].
La tesi che il problema sia quello di «'risolvere' i bias» grazie a interventi tecnici non è che un «seducente diversivo»[90], escogitato da aziende in conflitto di interessi che mirano a depoliticizzare la questione. Le denunce del carattere mistificatorio di tale narrazione sono perciò accompagnate dalla richiesta di finanziare ricerche indipendenti dalle Big Tech[91] e dalla proposta di concettualizzare la questione nei termini della tutela dei diritti umani[92].
Il modello di business delle grandi compagnie tecnologiche, fondato sull'appropriazione e la commercializzazione dei dati personali, sta infatti sfruttando una «bolla giuridica»[93]: ha cioè luogo in violazione di diritti giuridicamente tutelati, con la scommessa su un successivo salvataggio giuridico, in nome dell'inarrestabilità dell'innovazione tecnologica[94].
La priorità dei diritti individuali specificamente protetti dalla legge su un generico principio di innovazione[95], nonché l'evidenza delle violazioni di tali diritti, quando si utilizzino sistemi di apprendimento automatico per attività che hanno effetti rilevanti sulle vite delle persone, stanno a fondamento della proposta – formulata da Frank Pasquale e Gianclaudio Malgieri[96] – di una disciplina che preveda una presunzione di illegalità, ossia un regime di «illegalità di default»: fino a prova contraria, i sistemi di intelligenza artificiale ad alto rischio, incorporati in prodotti e servizi, dovrebbero essere considerati illegali, e l'onere della prova contraria dovrebbe incombere alle aziende.
Con ciò, l'usuale impostazione nei termini di un diritto delle persone alla spiegazione, in termini descrittivi, del perché una decisione fondata su sistemi di intelligenza artificiale sia stata assunta (a cui si oppone «la triplice barriera del segreto commerciale, degli accordi di non divulgazione e della complessità tecnica»[97]), sarebbe sostituita da un'automatica illegalità di tali decisioni, e dall'obbligo, per le aziende che volessero invece farne uso, di fornirne una giustificazione in termini normativi, ossia la dimostrazione che il processo di decisione automatizzato non è discriminatorio, non è manipolatorio, non è iniquo, non è inaccurato e non è illegittimo nelle sue basi giuridiche e nei suoi scopi.
Proposte analoghe sono fondate sull'analisi tecnica delle caratteristiche dei sistemi di apprendimento automatico: i sistemi di ottimizzazione predittiva dovrebbero essere proibiti *tout court*, nei casi in cui le decisioni abbiano conseguenze rilevanti sulle vite delle persone, poiché si fondano su false promesse[98]; per la medesima ragione, le narrazioni di chi ne sostenga, a fini commerciali, l'esistenza, sono da assimilarsi a pubblicità ingannevole.

---

[89] E. Harlan, O. Schnuck, *Objective or biased. On the questionable use of Artificial Intelligence for job applications*, February 16, 2021, https://interaktiv.br.de/ki-bewerbung/en/ .
[90] J. Powles, H. Nissenbaum, *The Seductive Diversion of 'Solving' Bias in Artificial Intelligence*, «OneZero», December 7, 2018, https://onezero.medium.com/890df5e5ef53.
[91] T. Gebru, *For truly ethical AI, its research must be independent from big tech*, «The Guardian», December 6, 2021, https://www.theguardian.com/commentisfree/2021/dec/06/google-silicon-valley-ai-timnit-gebru; D. Baker, A. Hanna, *AI Ethics Are in Danger. Funding Independent Research Could Help*, «Stanford Social Innovation Review», 2022, https://doi.org/10.48558/VCAT-NN16.
[92] V. Prabhakaran, M. Mitchell, T. Gebru, I. Gabriel, *A Human Rights-Based Approach to Responsible AI*, 2022, https://arxiv.org/abs/2210.02667.
[93] M. Giraudo, *Legal Bubbles*, in *Encyclopedia of Law and Economics*, a cura di A. Marciano, G.B. Ramello, New York, Springer, 2022, https://www.researchgate.net/publication/357702553.
[94] J. Stilgoe, *Who's Driving Innovation? New Technologies and the Collaborative State*, cit.
[95] Sul principio di innovazione, quale maschera dietro la quale grandi soggetti economici rivendicano la tutela dei loro concreti interessi, v. A. Saltelli, D.J. Dankel, M. Di Fiore, N. Holland, M. Pigeon, *Science, the endless frontier of regulatory capture*, cit., pp. 7-8.
[96] F. Pasquale e G. Malgieri, *From Transparency to Justification: Toward Ex Ante Accountability for AI*, «Brussels Privacy Hub Working Papers», VIII, 33, 2022, pp. 1-27, https://brusselsprivacyhub.com/wp-content/uploads/2022/05/BPH-Working-Paper-vol8-N33.pdf.
[97] Ivi, p. 2
[98] A. Wang, S. Kapoor, S. Barocas, A. Narayanan, *Against Predictive Optimization: On the Legitimacy of Decision-Making Algorithms that Optimize Predictive Accuracy*, cit.

Non si tratterebbe che di porre fine alla generale violazione di diritti giuridicamente tutelati: i sistemi di ottimizzazione predittiva impediscono infatti alle persone, al di fuori di quanto previsto dalla legge, di accedere a risorse o esercitare diritti. L'adozione di tali sistemi in ambiti quali l'assistenza sociale equivale – come ha osservato il Rapporteur Speciale delle Nazioni Unite per la povertà estrema e i diritti umani– alla decisione, in via amministrativa, di istituire delle «zone pressoché prive di diritti umani» («almost human rights-free zones»)[99]. Si può, ad esempio, essere inseriti in una *nofly list* – e non essere perciò ammessi a salire su un aereo – in virtù di un «sospetto» (ossia di una correlazione non specificabile) o vedersi negare la prescrizione di oppioidi, di cui pur si abbia un documentato bisogno, in quanto si è proprietari di cani ai quali sono stati prescritti farmaci simili – con ricette, com'è consuetudine, intestate ai proprietari – o in quanto si è stati vittime di abusi sessuali e si è per ciò stesso classificati come a rischio dal sistema di previsione algoritmica del rischio di overdose (Overdose Risk Score)[100].

In ambiti quali la giustizia, la salute, l'educazione o la finanza, nei quali si ha diritto a una spiegazione delle decisioni che ci riguardino, dovrebbe essere obbligatorio l'uso di sistemi che, a differenza di quelli di apprendimento automatico, siano fondati su modelli espliciti e su variabili interpretabili, e la costruzione di «catene di fornitura dei dati» che siano progettate, generate e curate, di volta in volta, in modo coerente con il sistema da costruire[101]. Tali operazioni, che il banale buon senso suggerirebbe, non sono tuttavia intraprese, in virtù del maggior costo che esse comporterebbero, rispetto alla cattura di quantità enormi di dati attraverso meccanismi di sorveglianza, e in virtù delle minori applicazioni commerciali di sistemi trasparenti, privi dell'aura magica della chiaroveggenza algoritmica. Le aziende scelgono perciò di includere, tra i costi da esternalizzare, quelli che derivano dai danni sociali prodotti dai sistemi di ottimizzazione predittiva.

Il contrasto non è dunque, in realtà, tra il rispetto dei diritti umani e il principio di innovazione, ma tra il rispetto dei diritti umani e il modello di business dei grandi monopoli del capitalismo intellettuale[102].

---

[99] P. Alston, *The Digital Welfare State – Report of the Special Rapporteur on Extreme Poverty and Human Rights*, 2019, https://daccess-ods.un.org/access.nsf/Get?OpenAgent&DS=A/74/493&Lang=E.

[100] D. McQuillan, *Resisting AI. An Anti-fascist Approach to Artificial Intelligence*, cit., pp. 76-78.

[101] N. Cristianini, *Shortcuts to Artificial Intelligence*, cit.; Idem, *La scorciatoia. Come le macchine sono diventate intelligenti senza pensare in modo umano*, cit.

[102] U. Pagano, *The Crisis of Intellectual Monopoly Capitalism*, «Cambridge Journal of Economics», XXXVIII, 2014, pp. 1409-1429, https://ssrn.com/abstract=2537972; T. Wu, *The Curse of Bigness. Antitrust in the New Gilded Age*, New York, Columbia Global Reports, 2018; M. Giraudo, *On legal bubbles: some thoughts on legal shockwaves at the core of the digital economy*, «Journal of Institutional Economics», XVIII, 4, 2022, pp. 587-604, https://doi.org/10.1017/S1744137421000473; S. Zuboff, *Surveillance Capitalism or Democracy? The Death Match of Institutional Orders and the Politics of Knowledge in Our Information Civilization*, «Organization Theory», III, 3, 2022, https://doi.org/10.1177/26317877221129290.