

FEMMINISTE col BOT

Tecnologie e intelligenze artificiali

DONNAWOMANFEMME DWF.IT DONNAWOMANFEMME

Redazione e amministrazione:
Via della Lungara, 19 – 00165 Roma

Redazione: Giada Bonu Rosenkranz, Patrizia Cacioli, Federica Castelli,
Noemi Ciarniello, Teresa Di Martino, Roberta Paoletti

Illustrazione e grafica: Alessandra Conti
E-mail: redazione@dwf.it
Sito Web: www.dwf.it

Contributo pubblicazioni per il 2024:
Italia: € 38,00
Paesi nell'area dell'Euro: € 50,00
Paesi extra-area dell'Euro: € 50,00
(o equivalente in dollari US)

I versamenti vanno effettuati:
sul c/c postale n. 000009134108 (IBAN IT50P076010320000009134108) intestato a
Associazione UTOPIA, via della Lungara, 19 - 00165 - Roma
Spediz. abbon. postale – DL 353/2003
(conv. in L. 27/02/2004 n° 46) art. 1 comma 1 – DCB Roma

Autorizzazione del Tribunale di Roma
n. 122 del 7 marzo 1986.

Finito di stampare nel mese di novembre 2024
Da CSR – Centro Stampa e Riproduzione Srl
Via di Salone 131/C, 00131 Roma

Trimestrale. Dir. Resp. Teresa Di Martino

©2024 . Editrice Associazione UTOPIA
Tutti i diritti sono riservati.

La riproduzione integrale o parziale dei testi
è vietata. Per eventuali utilizzazioni e traduzioni
richiedere l'autorizzazione alla direzione editoriale.

I files o dattiloscritti, anche se non pubblicati, non si restituiscono.

ISSN 0393-9014



DWF

2024, 2 N° 142

Sommario

editoriale

5 Nota editoriale

MATERIA

9 Genere e tecnologia: è questione di pratiche politiche
Mariacristina Sciannamblo

15 Intelligenza artificiale, *machine learning* e discriminazione.
L'AI alla prova con le nuove soggettività
Teresa Numerico

27 Come uno spillo nella bolla
Marta Capesciotti

37 Le cose si rompono. Donne* nella tecnologia che spaccano
computer e preconcetti
Aileen Dering

49 Anime eclettiche
lula

55 Un approccio femminista hacker all'intelligenza artificiale
Agnese Trocchi

67 Le questioni *vitali* dell'intelligenza artificiale e la tecnoscienza femminista
Federica Timeto

77 Dalla parte delle macchine
Ippolita

POLIEDRA

84 Manifest-No
Marika Cifor, Patricia Garcia, TL Cowan, Jasmine Rault, Tonia Sutherland, Anita Say Chan, Jennifer Rode, Anna Louren Hoffmann, Niloufar Salehi, Lisa Nakamura

93 Afro-Feminist Data Futures Manifesto
Pollicy

selecta

98 *Martina Micciché. Femminismo di periferia. Sonda, 2024*
Recensione di Serena Olcuire

100 *Paola Stelliferi. Tullia Romagnoli Caretoni nell'Italia repubblicana. Una biografia politica Viella, 2022*
Recensione di Eleonora Cirant

Intelligenza artificiale, *machine learning* e discriminazione. L'AI alla prova con le nuove soggettività

Teresa Numerico

Il test di Turing

“Intelligenza artificiale” è un termine nato da poco meno di settant’anni, nel 1956, come titolo di un seminario estivo al Dartmouth College, negli Stati Uniti. Il progetto disciplinare data forse qualche anno prima, dal progetto di Alan Turing di costruire più modestamente il concetto di *Machine intelligence*, un’intelligenza meccanica.

È interessante notare che nell’articolo del 1950 *Computing Machinery and Intel-*

ligence, Turing abbia inventato una sorta di gioco per dimostrare in che modo la macchina potrebbe essere considerata capace di pensare, senza bisogno di rispondere alle complesse domande intorno a cosa sia il dispositivo e, soprattutto, cosa sia il pensare. Il gioco, notissimo come Test di Turing, consisteva nell'imitarne uno che precedentemente veniva descritto come un gioco tra due persone di sesso diverso chiuse in stanze separate che avrebbero dovuto essere interrogate da un giudice inesperto e imparziale. Nella versione originaria del gioco, l'uomo avrebbe avuto il compito di mentire facendo credere di essere una donna, mentre la donna avrebbe dovuto dire la verità, cercando di convincere l'interrogante.

L'obiettivo dell'imitazione era sostituire la donna con la macchina e chiedere alla macchina di fare finta di comportarsi come un essere umano, traendo così in inganno il giudice inesperto dopo cinque minuti di conversazione su qualsiasi argomento.

Nonostante questo particolare sia stato a lungo ignorato dall'innumerevole letteratura critica sull'argomento, esso è stato finalmente messo in luce (cfr. Genova 1994), anche tenendo conto della sofferenza di Turing stesso, stretto dall'identificazione cisgender e considerate le sue inclinazioni omosessuali, che gli costarono una condanna a subire un trattamento ormonale – la castrazione chimica – il quale forse contribuì a condurlo al suicidio nel 1954. Tuttavia, una più recente interpretazione intercetta una questione più sottile rispetto al Test e alla formulazione a partire dalla differenza uomo-donna, cioè la dimensione di tutto o nulla, acceso/spento, sì/no relativa alla questione di genere e quindi anche attribuita all'intelligenza della macchina (cfr. Kind 2022).

Kind parte dal presupposto che lo stesso Turing, nel corso della descrizione del suo test, lo definisca come un dispositivo che cerca attraverso la finzione di una conversazione di simulare un essere umano, senza ulteriori specificazioni di genere.

La partenza del gioco, quindi, ci suggerirebbe non tanto una declinazione di genere del progetto quanto questo aspetto classificatorio binario che sarebbe al centro dello sviluppo dell'intelligenza meccanica: l'idea che si possa dire se la macchina sia intelligente o se non lo sia senza ulteriori sfumature. È questo il nodo teorico che provo a esplorare in questo contributo, mostrando come il sistema di interpretazione algoritmica dei dati proprio del *machine learning* ci conduca direttamente a una classificazione binaria dei soggetti e delle loro caratteristiche che mal si adatta alla complessità delle relazioni sociali e psichiche delle soggettività dinamiche.

Intelligenza artificiale tra reasoning e learning

Ho fatto riferimento a questa interessante interpretazione del test di Turing perché uno dei principali obiettivi di questa nuova fase dell'intelligenza artificiale, cominciata pressappoco nel 2012, è concentrata su progetti di carattere classificatorio e predittivo.

L'etichetta "intelligenza artificiale" è un termine ombrello per un'ampia serie di metodi sviluppati fin dagli anni Cinquanta. Possiamo distinguerli in due macrocategorie: quelli basati sul ragionamento e quelli basati sull'apprendimento. I metodi basati sul ragionamento sono stati i primi in ordine di tempo a essere sperimentati, e prevedevano la definizione delle regole attraverso le quali le macchine avrebbero dovuto esercitare la propria capacità inferenziale. Questo processo si considerava almeno in parte erede della logica e si riprometteva di simulare le funzioni alte degli esseri umani, come la capacità di risolvere problemi, dimostrare teoremi o pianificare compiti complessi. I metodi simulavano quindi le funzioni mentali degli esseri umani e si basavano su un modello inferenziale di tipo deduttivo che, partendo da premesse prestabilite, ne inferivano tutte le conseguenze. Nonostante le grandi aspettative e gli ingenti investimenti, questi sistemi non hanno prodotto i risultati attesi e hanno progressivamente perso la battaglia retorica rispetto alla costruzione di macchine realmente in grado di sostituire gli esseri umani nei compiti considerati intelligenti.

Dagli anni Ottanta in poi hanno avuto successo quelle tecnologie basate sull'apprendimento che non miravano a replicare il ragionamento umano, ma piuttosto a simulare l'organizzazione delle reti neurali cerebrali, sebbene in modo molto stilizzato. Questi metodi erano già stati inventati fin dalle origini, ma all'inizio non avevano avuto abbastanza finanziamenti e, soprattutto, non erano disponibili le infrastrutture di calcolo necessarie al loro successo, né era facile costruire delle basi di dati sufficientemente grandi da poterle addestrare adeguatamente.

La caratteristica dei sistemi di apprendimento automatico è che, diversamente dal modello del *reasoning*, non fanno riferimento a un insieme di ragionamento prestabilito, ma estraggono le regolarità attraverso sistemi di interpretazione algoritmica che lavorano sui dati analizzati. Dal momento in cui si sono diffuse le piattaforme digitali è stato possibile raccogliere grandi quantità di dati relativi ai comportamenti, alle preferenze e alle relazioni tra le persone in formato direttamente leggibile dalla macchina. La grande disponibilità dei dati ha permesso di addestrare gli algoritmi di apprendimento a riconoscere gli schemi dentro i dati del passato per proiettarli sul futuro, mettendo i nuovi sistemi in condizione di fare previsioni, usate per fornire raccomandazioni commerciali, ma anche per altri scopi più impegnativi e rilevanti per prendere decisioni che avevano delle conseguenze non sempre positive sulla vita delle persone.

Apprendimento e discriminazione

Il mantra dei sistemi basati sull'apprendimento consiste nel rivendicare la mancanza di una spiegazione per i sistemi di classificazione e per le giustificazioni alle previsioni future. La quantità di dati permetterebbe di sopperire all'assenza di un modello dei fenomeni sotto osservazione. La tesi prevalente, suggerita da un famoso articolo su *Wired* del 2008 dal titolo *The end of theory*, scritto da Chris Anderson, è che i modelli di spiegazione sono comunque sbagliati e sempre più spesso è possibile evitare di usarli grazie alle grandi quantità di informazioni disponibili. Tuttavia, queste affermazioni sono troppo superficiali ed epistemicamente poco sostenibili: sono molti i casi nei quali si è potuto dimostrare sul piano del controllo dei risultati che questi strumenti basati sull'apprendimento siano facilmente spinti a compiere inferenze scorrette e discriminatorie.

Un caso eclatante è quello di un sistema progettato da Amazon per la selezione del personale addestrato sui dati dei dipendenti dell'azienda che automaticamente scartava i curricula delle donne, perché non corrispondevano al modello estratto dai dati sul candidato ideale (Dastin 2018). Il progetto fece così tanto scalpore che non si andò mai oltre la fase di test.

In molti casi i sistemi di *machine learning* dedicati a valutazioni e previsioni hanno prodotto esiti discriminatori. Nel 2018 Joy Buolamwini e Timnit Gebru, nel loro articolo *Gender shades*, hanno dimostrato che la maggior parte dei sistemi di riconoscimento facciale allora in uso non sapessero riconoscere le donne afroamericane, manifestando una discriminazione intersezionale relativa ad alcune soggettività marginali, assenti o scarse nel database di addestramento.

Un altro più recente esempio del potenziale discriminatorio degli algoritmi addestrati sui dati riguarda un'analisi svolta sugli algoritmi che presiedono all'intelligenza artificiale generativa, *text-to-image*, nel quale si dimostra che valori e pregiudizi sono incorporati nella base dati dei meccanismi di addestramento e restituiscono una rappresentazione discriminatoria e scorretta del mondo, inclusa la sua dimensione immaginaria. Per esempio, si dimostra che la rappresentazione della bellezza femminile riproduce lo stereotipo della donna giovane, bianca, bionda e cerulea (Luccioni et al. 2023a, 2023b).

Molte studiosi che si occupano di rapporti tra femminismo e tecnologia hanno sostenuto che affidarsi ai dati disponibili senza controllarli rende molto probabile reperire correlazioni che costruiscano una rappresentazione parziale e discriminatoria delle donne e delle altre soggettività marginali o fuori norma; tra loro spiccano Catherine D'Ignazio e Lauren Klein, autrici di *Data Feminism* (2023). Tale effetto è causato dalla ricorrenza di informazioni discriminatorie nei data set di addestramento che poi intervengono a definire negativamente

te le soggettività più deboli, a causa della mancanza di corrispondenza con gli standard prescritti. I sistemi di traduzione automatica, per esempio, associano in modo prevalente i ruoli inferiori alle donne e quelli professionalmente più qualificati agli uomini, passando da una lingua senza generi a una che ce li ha: infermiere è donna e dottore è uomo. Questo effetto è il prodotto delle ricorrenze che si trovano nei dati, i quali riflettono l'orientamento del mondo, che a sua volta non è ideale: vi si consumano piccole e grandi ingiustizie riprodotte nei sistemi di apprendimento. I dati, che non provengono dal laboratorio e non sono raccolti in un contesto sterile, raccolgono le sperequazioni di genere, origine e classe, talvolta amplificandole.

Alcuni esperti suggeriscono che basterebbe emendare i dati per limitarne gli effetti distorsivi, ma questo finirebbe per rendere i sistemi ancora più arbitrari e potrebbe anche condurre a rimedi peggiori dei mali. L'incidente avvenuto a Gemini, l'intelligenza generativa di Google, nel febbraio 2024, dimostra quanto ritoccare i dati senza controllo possa produrre esiti indesiderati e pericolosi. Nel tentativo di evitare le distorsioni, si rappresentavano i padri fondatori americani come donne nere e i guerrieri dell'antica Grecia come uomini e donne asiatiche (Robertson 2024). Inoltre, se si manipolano i dati, non siamo più di fronte a un sistema algoritmico automatico, e dovrebbe essere esplicito quali siano stati gli interventi e in quale direzione si siano orientati: sono apparati socio-tecnici, ma sarebbe importante evidenziare quale contributo umano di "aggiustamento" sia stato inserito, per evitare altre potenziali criticità inattese.

Un'altra questione che andrebbe approfondita è l'ampiezza temporale dei dati. Riteniamo questi dati completi, ma appartengono spesso soltanto al tempo in cui è stato possibile raccogliergli direttamente in formato digitale, e quindi sono inevitabilmente affetti da presentismo: non sappiamo se e come sarebbero le interpretazioni algoritmiche qualora i dati fossero più estesi storicamente. Certamente usare un metodo come l'induzione come principale strumento di interpretazione dei dati non permette di riconoscere le rotture epistemiche e le conseguenti soluzioni di continuità adottate di volta in volta. Tutto questo è radicalmente espulso da quei sistemi di comprensione. E non a caso, perché si tratta di processi che strutturalmente vogliono disciplinare il mondo secondo le proprie categorie e orientarlo, non comprenderlo nelle sue sfumature complesse. L'obiettivo della previsione cancella l'interesse per la riflessione e la spiegazione dei fenomeni sociali e soggettivi: il drive è esclusivamente concentrato sull'intervento, che ha spesso un carattere di disciplinamento.

Il potenziale di rischio nell'equivocare il senso dei fenomeni è spesso anche legato a quali dati si raccolgono e con quale intento, ma per coloro il cui intento è governare le preferenze, le azioni e le relazioni, non ha molto valore minimizzare aspetti importanti della realtà. Tale riflessione critica si concretizza in una

parte delle femministe che si interessano di questioni tecnologiche nell'attivare ricerche per costruire pratiche di raccolte alternative dei dati. Cito in questo campo alcune studiose come Paula Ricaurte Quijano e Catherine D'Ignazio, che lavorano per una raccolta critica, creativa e politica dei dati. Per esempio, registrare i dati dei femminicidi in sud America significa anche risignificare questo evento. Molte uccisioni di donne, infatti, non vengono rubricate come femminicidi da parte dei loro mariti, ex mariti o amanti, e tale meccanismo conduce a escluderli dalle statistiche. Catherine D'Ignazio, con il libro *Counting Feminicide: Data Feminism in Action* (2024), formula una strategia per riprendere il controllo dei dati dai quali estrarre il senso in modo critico, e non come una semplice proiezione del passato sul futuro. Al contempo, Paula Ricaurte (2023) studia come sia possibile resistere alla visione algoritmica dei corpi/luoghi da parte delle comunità indigene di Abya Yala.

Eppure, non tutto si può giocare esclusivamente sui dati, sebbene curarli e avere chiaro lo sguardo e la prospettiva della raccolta abbia indubitabili vantaggi.

L'inevitabile incertezza di ogni previsione

Al di là dei tanti esempi concreti che dimostrano l'inefficacia delle inferenze algoritmiche nell'interpretazione dei dati rispetto al rischio dell'infiltrarsi di giudizi discriminatori nelle pratiche di valutazione, possiamo osservare una questione più profonda di carattere epistemico che mostra inequivocabilmente l'orientamento verso standard penalizzanti per le soggettività marginali, che hanno un carattere strutturalmente intersezionale.

Considerare le previsioni come l'obiettivo degli sforzi algoritmici interpretativi, in assenza di un modello di spiegazione nel quale la proiezione futura si possa inquadrare, produce inevitabilmente uno strappo epistemico. Il futuro è sempre incerto soprattutto nei contesti contingenti come quelli della vita umana. Sostenere di poter estrarre le regolarità dai dati passati per proiettarle nel futuro senza ulteriori precisazioni significa utilizzare l'inferenza induttiva che presuppone la regolarità del contesto, cioè prescrive il mondo stabile (Gigerenzer 2022). Nei contesti sociali, dove la contingenza è la norma non abbiamo nessuna garanzia che il passato corrisponda al futuro. Nonostante le previsioni si possano rivelare esatte in tante circostanze, ciò non ne garantisce l'affidabilità: potremmo infatti incontrare un caso in cui la previsione fallisce anche dopo milioni di previsioni corrette, senza contare il carattere prescrittivo implicito nella previsione di una soggettività considerata molto autorevole (Rouvroy, Berns 2013).

Wendy Chun nel suo *Discriminating data* (2021) introduce il concetto di "omofilia" per descrivere scelte, preferenze e comportamenti delle persone; secondo

questo principio chi vive nei ghetti non sarebbe costretto a farlo, ma seguirebbe la predisposizione a vivere coi propri simili. Tale approccio, secondo Chun, non tiene minimamente conto delle condizioni economiche di partenza, o delle scelte discriminatorie degli affittuari. La tesi dell'omofilia naturalizza le situazioni di esplicita segregazione, con la complicità di una tendenza immaginaria degli attori sociali ad agire sempre liberamente.

I criteri e i meccanismi che presiedono ai processi inferenziali attivi nei sistemi di intelligenza artificiale attualmente in funzione sono orientati a scopi precisi di massimizzazione dei profitti, e definiscono la rappresentazione del mondo sociale e, soprattutto, la sua previsione in modi che consentano la manipolazione; questo è il caso dei sistemi di raccomandazione, usati dalle piattaforme di *social networking*, di commercio online e di *streaming*. Possiamo immaginare quanto possa essere pericoloso trasportare le inferenze algoritmiche usate nei contesti di *marketing* ad altre aree della presa di decisione, dove la loro valutazione incerta potrebbe avere conseguenze molto più pericolose del ricevere proposte di acquisto alle quali non siamo inclini.

Non sarebbe impossibile usare gli algoritmi di raccomandazione e la conseguente attività di decisione o di intervento sulle persone in modo da evitare le discriminazioni e realizzare azioni positive per riequilibrare le società contro effetti intersezionalmente discriminatori, ma non è questo l'obiettivo che si propongono i grandi sistemi delle aziende tecnologiche.

Possiamo individuare a questo proposito alcuni punti critici che sottendono al funzionamento della predizione algoritmica basata su criteri probabilistici riferiti all'interpretazione dei dati del passato: proiettare il passato sul futuro significa presumere che le situazioni non cambino, e che se qualcuno ha commesso uno sbaglio o ha una cattiva abitudine continuerà ad libitum, e soprattutto inchioda le persone alla loro rappresentazione usuale e condivisa, che è un metodo sicuro per preservare le discriminazioni e i pregiudizi. Esso prescrive, inoltre, che il contesto sociale e materiale dal quale emergono queste caratteristiche sia stabile e immutabile e non sia il frutto di processi storici e possa modificarsi nel tempo – una sorta di cattura del futuro da parte del passato. Considerare lo *status quo* come inevitabile giustifica una visione conservatrice della società. Fotografare il presente come immutabile è una pratica parziale e omissiva, ma è irriducibilmente usata nel *machine learning*.

Questo esito è particolarmente pericoloso e lesivo di tutti i soggetti deboli che hanno una storia di discriminazione dalla quale la loro identità viene plasmata in maniera intersezionale: tutte le marginalità che provengono dalle condizioni di povertà, di origine etnica, di genere, o di non conformità alle categorie che sono definite come standard sono danneggiate da un sistema inferenziale che

ciecamente rilancia in maniera probabilistica le condizioni e le abitudini del passato sul futuro. La discriminazione in questo caso è doppia, perché naturalizza la storia di discriminazioni e stereotipi facendola apparire come l'esito di una valutazione tecnica e neutrale, e rendendo, quindi, impossibile scrollarsi di dosso le condizioni di iniquità provenienti dalla propria storia individuale e collettiva.

Discrezione e decisione: alcune osservazioni finali

Una questione sulla quale spesso si discute rispetto all'intelligenza artificiale riguarda ciò che può essere automatizzato: per attivare la delega alla macchina dobbiamo decidere se effettivamente possiamo fidarci della sua valutazione nei contesti in cui vogliamo appoggiarci ai sistemi tecnici.

Da diversi risultati anche di carattere socio-economico (Sen, 2014) sappiamo, infatti, che nella soluzione ai problemi multicriterio – come tutti i problemi dove le decisioni sono aleatorie e impattano sulle vite delle persone, e coinvolgono interessi distinti e magari contrapposti – non è possibile adottare una soluzione puramente razionale, ma occorre negoziare politicamente la soluzione. Pensiamo alle scelte di reclutamento, all'accesso al credito, al welfare, alle cure mediche, alla valutazione della pericolosità dei e delle responsabili per un reato. In questi casi è strutturalmente impossibile avere una soluzione ottima e, quindi, usare esclusivamente la razionalità per scegliere. Tale consapevolezza implica che la distinzione tra regole per automatizzare il processo di valutazione e discrezione non siano veramente contrapposti. La lezione di Wittgenstein (1953) ci insegna che seguire una regola è più complicato di quanto crediamo e soprattutto prevede l'intervento di un certo potere di scelta e di molta interpretazione.

Nel libro *Rules*, Lorraine Daston mostra come sia sempre possibile rintracciare la discrezionalità in ogni processo algoritmico di decisione. Questa discussione prova che gli strumenti di machine learning per calcolare probabilisticamente la soluzione dei nostri roveli decisionali sono, comunque, intessute di discrezionalità a ogni livello: a livello dei dati di addestramento che servono per costruire i pattern e le regolarità da seguire, a livello di meccanismi algoritmici che definiscono le regolarità di cui tenere traccia, al livello di spazi latenti costruiti dagli strumenti di AI generativa, al livello delle decisioni prese dalle molte persone che partecipano alle fasi di addestramento, e soprattutto al livello di funzione obiettivo che deve sempre essere definita. Per decidere se un algoritmo produca un risultato soddisfacente, ho bisogno di una funzione di valutazione dell'output; ma nei casi dove regna l'incertezza, come le situazioni contingenti, la definizione di quella funzione instaura un criterio a cui l'algoritmo si atterrà.

Tale criterio non è sempre esplicito nemmeno a chi lo definisce, ma è cogente rispetto alla discrezionalità che esercita.

La valutazione probabilistica, che è il classico risultato di questi sistemi, non è mai valutabile nel giudizio singolo, ma solo nella serie di valutazioni, e a condizione di avere la possibilità di confrontare l'adeguatezza delle previsioni con come siano andate le cose concretamente. Nessun giudizio probabilistico della forma: "la candidata è al 90% inadatta alla posizione" è falsificabile sul caso singolo perché in questo giudizio è compreso l'opposto "al 10% la candidata è adeguata alla posizione". Si tratta, quindi, di una valutazione oracolare che, come sappiamo, non può mai venire falsificata. Ma, proprio per questo, non garantisce nessun carattere di scientificità alla valutazione espressa.

Per rendere più eque le decisioni supportate dai sistemi tecnici occorrerebbe accoglierne la discrezionalità e farla emergere, in modo da valutare la sua adeguatezza al contesto e alle singolarità delle situazioni che ne sono oggetto, consentendo a chi si dovesse sentire discriminato la possibilità di accesso ai criteri e all'orientamento che ha prodotto la scelta. Le regole dovrebbero essere espresse in modo esplicito e rivelare la posizione dalla quale si guarda ai casi concreti: tutte caratteristiche che al momento i sistemi di intelligenza artificiale non garantiscono quando esercitano procedure predittive che preludono alla decisione, siano esse relative al supporto alla deliberazione, alla generazione o al riconoscimento di contenuti.

Teresa Numerico

Teresa Numerico è professoressa associata di Logica e Filosofia della Scienza all'Università di Roma Tre. Si occupa di filosofia della tecnologia, politica dell'intelligenza artificiale, nuova epistemologia digitale delle scienze umane e sociali. Pubblica regolarmente su riviste internazionali. Tra i suoi libri: *Alan Turing e l'intelligenza delle macchine* (FrancoAngeli, 2005), *Web Dragons* (con M. Gori e I. Witten; Morgan Kaufmann, 2007), *L'umanista digitale* (con D. Fiorimonte e F. Tomasi; Il Mulino, 2010; trad. ingl. aggiornata, *The digital humanist*; Punctum Books, 2016) e *Big Data e Algoritmi* (Carocci, 2021). Ha una figlia adolescente, Maia, che considera il suo successo più straordinario.

Riferimenti bibliografici

- ANDERSON C., (2008), *The end of theory The data deluge makes the scientific method obsolete in Wired magazine*, 16(7).
- BUOLAMWINI J., GEBRU T., (2018, January), *Gender shades Intersectional accuracy disparities in commercial gender classification*. In *Conference on fairness, accountability and transparency*, PMLR, pp. 77-91.
- CHUN W. H. K., (2021), *Discriminating data Correlation, neighborhoods, and the new politics of recognition*, MIT press, Cambridge, Ma.
- D'IGNAZIO C., KLEIN L. F., (2023), *Data feminism*, MIT press, Cambridge, Ma.
- D'IGNAZIO C., (2024). *Counting femicide Data feminism in action*. MIT PRESS, CAMBRIDGE, MA.
- DASTIN J., (2018), *Insight - Amazon scraps secret AI recruiting tool that showed bias against women*, Reuters, (11/10/2018), <https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/>
- DASTON L., (2022), *Rules*, Princeton University Press, Princeton, Oxford.
- GENOVA J., (1994), *Turing's Sexual Guessing Game*, in *Social Epistemology* 8(4), pp. 313– 326.
- GIGERENZER G., (2022), *How to Stay Smart in a Smart World*, Mit Press, Cambridge (Mass.); trad it. *Perché l'intelligenza umana batte ancora gli algoritmi*, Raffaello Cortina, Milano, 2023.
- KIND A., (2022) *Computing Machinery and Sexual Difference*, in Maitra K., McWeeny J. (eds) *Feminist Philosophy of Mind*, Oxford University Press, Oxford, pp. 54-69.
- LUCCIONI A.S., et al., (26/06/2023a) *Ethics and Society Newsletter #4 Bias in Text-to-Image Models*, <https://huggingface.co/blog/ethics-soc-4>
- LUCCIONI A. S., AKIKI C., MITCHELL M., JERNITE Y., (2023b), *Stable bias Analyzing societal representations in diffusion models*, arXiv preprint arXiv:2303.11408.
- RICAURTE P., (2023). Resistencia como reexistencia: la defensa del cuerpo-territorio en la sociedad algorítmica. *Pléyade*, 32, 64-92.
- ROBERTSON A., (2024), *Google apologizes for 'missing the mark' after Gemini generated racially diverse Nazis*, in *The Verge*, 21 Feb. 2024, <https://www.theverge.com/2024/2/21/24079371/google-ai-gemini-generative-inaccurate-historical>
- ROUVROY A., BERNS T., (2013). *Gouvernementalité algorithmique et perspectives d'émancipation*. Réseaux, 177(1), 163-196.
- SEN A., (2014), *Arrow and the impossibility theorem*. In Maskin E., Sen A., (Eds) *The Arrow impossibility theorem*, Columbia University Press, New York, pp. 29-42.
- TURING A.M., (1950) *Computing Machinery and Intelligence*, *Mind* 59 (236), pp.433-460.
- WITTGENSTEIN L., (1953), *Philosophische Untersuchungen, Philosophical investigations*, a Eds. G. E. M. Anscombe e Rush Rhees, Blackwell, Oxford 1953, 1958; *Ricerche filosofiche*, tr. it. di Piovesan R., Trincherò M., a cura di Trincherò M., Einaudi, Torino, 1967.