



Data di ricezione: 14.05.25 / Data accettazione: 12.08.25 / Data di pubblicazione: 31.12.25
doi: 10.82015/NNR.2025.100113

Dinamiche di potere, rapporti di genere e gestione del rischio nell'era dell'intelligenza artificiale

Power Dynamics, Gender Relations, and Risk Management in the Age of Artificial Intelligence

Michele Di Salvo*

Sintesi

L'articolo esplora la triade concettuale di genere, potere e rischio nell'ecosistema socio-tecnico dell'IA, argomentando che i sistemi algoritmici non sono neutrali, bensì agenti attivi nel plasmare le dinamiche di potere esistenti. Attraverso un'analisi critica che spazia dalla filosofia della tecnica alla sociologia dei media e agli studi femministi, si esamineranno i molteplici rischi dell'IA di genere: dai bias discriminatori incorporati nei dataset e negli algoritmi, alla sorveglianza differenziata e all'erosione dell'agency, fino alla riproposizione di stereotipi nei modelli generativi. Parallelamente, si indagheranno le asimmetrie di potere nella progettazione, sviluppo e governance dell'IA, dominata da una cultura tecnologica maschile e occidentale. L'articolo propone anche un percorso etico-politico per un'IA femminista, intersezionale e plurale, fondata su principi di giustizia algoritmica, partecipazione democratica e design responsabile. Con oltre 60 riferimenti bibliografici, il saggio rappresenta una mappatura critica e una bussola per orientarsi in uno dei territori più controversi e decisivi del nostro tempo.

* Direttore di Neural Nexus Review di Fondazione E.N.I.A. (Ente Nazionale per l'Intelligenza Artificiale) e Presidente del Consiglio di Amministrazione di CrossMedia Labs. È membro di numerose associazioni e gruppi di ricerca scientifica, tra cui la Society for Neuroscience, la European Federation of Neuroscience Societies, l'International Neuropsychoanalysis Society, la Cognitive Neuroscience Society e la International Neuroethics Society. La sua ricerca si concentra sui fondamenti neuroscientifici della psicologia. Orcid <https://orcid.org/0000-0002-9531-0591> / E-mail: mik.disalvo@gmail.com.



Parole Chiave: Intelligenza Artificiale; Genere; Bias Algoritmico; Femminismo Tecnologico; Sorveglianza; Potere; Rischio; Etica; Intersezionalità; Giustizia Sociale.

Abstract

The article explores the conceptual triad of gender, power, and risk within the socio-technical ecosystem of AI, arguing that algorithmic systems are not neutral but are active agents in shaping existing power dynamics. Through a critical analysis spanning from the philosophy of technology to media sociology and feminist studies, it will examine the multiple risks of gendered AI: from discriminatory biases embedded in datasets and algorithms, to differentiated surveillance and the erosion of agency, to the reproduction of stereotypes in generative models. Concurrently, the article will investigate the power asymmetries in the design, development, and governance of AI, dominated by a male and Western-centric technological culture. The article also proposes an ethical-political path towards a feminist, intersectional, and pluralistic AI, founded on principles of algorithmic justice, democratic participation, and responsible design. With over 60 bibliographic references, the essay represents a critical mapping and a compass for orientation to navigate one of the most controversial and decisive territories of our time.

Keywords: Artificial Intelligence; Gender; Algorithmic Bias; Technological Feminism; Surveillance; Power; Risk; Ethics; Intersectionality; Social Justice.



1. Introduzione: il mitòs e il lògos dell'intelligenza artificiale

C'è una narrazione dominante sull'Intelligenza Artificiale (IA), un mitòs potente che la dipinge come una forza storica, obiettiva e puramente razionale, un lògos superiore che finalmente libererà l'umano dall'errore e dal pregiudizio. È una narrazione seducente, che promette efficienza, ottimizzazione e una forma di giustizia matematica, incorruttibile dalle passioni umane. Eppure, questa promessa di neutralità si infrange quotidianamente contro l'evidenza dei fatti: algoritmi di reclutamento che penalizzano i curricula con parole "femminili", sistemi di riconoscimento facciale che falliscono nel riconoscere i volti delle donne di colore, assistenti vocali intelligenti a cui viene data una voce e una personalità femminile e sottomessa.

Come scrive la studiosa Kate Crawford nel suo fondamentale *Atlas of AI*, «l'IA non è un oggetto nel mondo, né è un costrutto puramente tecnico. È un'impresa industriale che richiede un'enorme quantità di risorse materiali e di lavoro» (Crawford, 2021, p. 8). Queste risorse – dai minerali estratti in condizioni spesso disumane, ai dati raccolti su scala planetaria, alla forza lavoro di etichettatura – non sono neutre. Sono immerse in relazioni di potere geopolitiche, economiche e sociali preesistenti. L'IA, in altre parole, è un'architettura del potere.

In questo articolo, ci addentreremo proprio nelle intricate relazioni tra questa architettura e la categoria del genere. Il genere non è un dato biologico, ma un costrutto sociale e performativo (Butler, 1990), un sistema di significati e norme che organizza la vita sociale, assegna ruoli, distribuisce risorse e struttura le identità. È, per sua natura, una dimensione del potere. L'IA, in quanto tecnologia che classifica, predice, ottimizza e governa, interagisce costantemente con questo sistema. Può cristallizzarlo, rendendo automatiche e invisibili le sue gerarchie; può amplificarlo, dando a pregiudizi impliciti una scala e una velocità inedite; o, potenzialmente, potrebbe sovvertirlo, se progettata con intenzione politica e consapevolezza critica.

Il rischio è l'altra faccia del potere in questo contesto. Non si tratta solo del rischio di un'ingiustizia puntuale, ma di un rischio sistemico: il rischio di creare un mondo in cui le



disuguaglianze di genere diventano così incorporate nelle infrastrutture tecnologiche da risultare naturali, inevitabili e quindi immuni alla contestazione politica. È il rischio di una nuova, subdola forma di determinismo tecnologico che, sotto le spoglie dell'oggettività, reifica le norme sociali dominanti.

Attraverso un percorso articolato, esploreremo come il genere sia codificato negli algoritmi, come l'IA agisca come strumento di sorveglianza e controllo dei corpi, e come i modelli linguistici e generativi riproducano stereotipi culturali. Analizzeremo poi le asimmetrie di potere nell'industria tech e le derive distopiche degli stati sorveglianti. Infine, cercheremo di tracciare una rotta alternativa, esplorando i principi per un'IA femminista e intersezionale e le vie concrete per una governance democratica e un design responsabile. L'obiettivo è fornire una bussola critica per navigare in un paesaggio in cui il futuro delle relazioni di genere è, letteralmente, in fase di codifica.

2. L'algoritmo sessista: come il pregiudizio diventa codice

Il punto di partenza obbligato per qualsiasi discorso su IA e genere è il problema dei bias algoritmici. Il termine "bias" rischia però di essere fuorviante, suggerendo una piccola deviazione, un'imperfezione tecnica in un sistema altrimenti perfetto. In realtà, come dimostrano innumerevoli casi di studio, questi bias sono spesso il riflesso fedele di pregiudizi strutturali della società.

2.1. I dati sono la nuova argilla: stereotipi e under-rappresentation

Il mantra del capitalismo digitale è che i dati siano il nuovo petrolio, la risorsa fondamentale per il complessivo sistema dei processi economici, dei servizi pubblici e dell'attività politico-amministrativa. Ma i dati non sono una materia prima grezza e pura; sono un prodotto culturale. Raccolgono le tracce digitali delle nostre società, con tutte le loro storture storiche, per addestrare gli algoritmi. Se un dataset utilizzato per addestrare un algoritmo fosse prevalentemente composto da volti maschili bianchi,



l'algoritmo interpreterebbe il "volto umano" come qualcosa che per definizione trova le proprie caratteristiche fondamentali nel genere maschile e nel colore bianco della pelle. Il celebre studio di Joy Buolamwini del MIT Media Lab, "Gender Shades", ha dimostrato proprio questo: i sistemi di riconoscimento facciale di aziende come IBM, Microsoft e Megvii avevano tassi di errore minimi per i volti di uomini bianchi, ma questi tassi schizzavano fino a quasi il 35% per i volti di donne di colore (Buolamwini e Gebru, 2018). L'algoritmo non era "rotto"; stava semplicemente rispecchiando, e quindi automatizzando, la sotto-rappresentazione e l'invisibilità delle donne nere nell'immaginario visivo dominante e, di conseguenza, nei dataset.

Lo stesso principio si applica al genere. I dataset linguistici su cui si addestrano i grandi modelli come GPT (il predecessore dei modelli che alimentano ChatGPT) sono estratti da vastissime porzioni di internet: Wikipedia, social media, siti di news, forum. Questi spazi non sono neutrali. Sono storicamente dominati da voci maschili e permeati da stereotipi di genere. Uno studio sull'enorme dataset "Common Crawl" ha rilevato una significativa sovrarappresentazione di parole e contesti associati a ruoli maschili e una correlazione tra nomi femminili e termini legati alla famiglia o all'aspetto fisico (Bolukbasi et al., 2016). Quando un algoritmo impara da questi dati, impara non solo la grammatica, ma anche i pregiudizi in essi contenuti.

2.2. La profezia che si autoavvera: gli algoritmi di reclutamento

Uno degli esempi più noti di bias di genere associati all'IA ci è stato fornito dall'algoritmo per le operazioni di reclutamento della forza lavoro adottato da Amazon. Per automatizzare la ricerca di talenti, l'azienda ha sviluppato uno strumento di machine learning a cui affidare la valutazione dei curricula. Il sistema fu addestrato sulle candidature ricevute dall'azienda nel decennio precedente, la stragrande maggioranza delle quali proveniva da uomini, riflettendo la dominanza maschile nel settore tech. L'algoritmo imparò quindi a penalizzare qualsiasi cosa associasse al "femminile": la parola "donne" (come in "club di scacchi femminile"), i nomi di college storicamente



femminili, e persino verbi e costruzioni linguistiche più comuni nei curriculum di donne (Dastin, 2018). In sostanza, l'algoritmo non stava cercando il "miglior candidato"; stava cercando il candidato che più assomigliava a quelli assunti in passato, cristallizzando così lo status quo e rendendo il sistema di assunzione una profezia che si autoavvera. Questo caso è paradigmatico perché mostra come l'IA non faccia che replicare i pattern del passato, ma presentandoli come scelte "ottimali". L'efficienza dell'algoritmo diventa uno strumento di conservazione del potere, non di sua trasformazione. Il rischio qui è la naturalizzazione della discriminazione: se un sistema "oggettivo" sceglie un uomo invece di una donna, la colpa non può più essere imputata a un recruiter sessista, ma a una logica matematica inappellabile. La discriminazione diventa opaca, tecnicizzata e quindi più difficile da identificare e contestare.

2.3. La finzione della neutralità e il feticcio dell'oggettività

Alla base di questi fenomeni c'è quella che Safiya Umoja Noble (2018) definisce, nel suo libro *Algorithms of Oppression*, la «*finzione della neutralità tecnologica*» (Noble, 2018). L'ideologia della Silicon Valley ha a lungo propagandato l'idea che il codice sia al di sopra della politica, della razza e del genere. Eppure, ogni fase del processo di sviluppo dell'IA – dalla scelta del problema da risolvere, alla raccolta dei dati, alla definizione delle metriche di successo – è intrisa di giudizi di valore umani.

La scelta di automatizzare il reclutamento, piuttosto che, ad esempio, lo screening per la parità salariale, è già di per sé una scelta politica. La definizione di "successo" per un algoritmo (ad esempio, massimizzare i clic o minimizzare i costi) incorpora valori specifici che spesso non sono allineati con la giustizia sociale. Come sostiene la filosofa Shannon Vallor (2016), l'etica non è un qualcosa che si può "aggiungere" a un sistema già sviluppato; deve essere integrata nel processo di progettazione fin dall'inizio, attraverso quella che lei chiama *formazione tecno-morale* (Vallor, 2016).

L'oggettività, in questo contesto, diventa un feticcio. Si confonde l'assenza di emozioni umane con una superiore forma di verità. Ma la "decisione" algoritmica non è altro che il



risultato di complesse operazioni matematiche su dati che sono, per loro natura, parziali e situati. Affidarsi ciecamente a questi sistemi significa quindi abdicare alla propria capacità di giudizio in favore di un'autorità tecnocratica che, dietro la maschera della neutralità, nasconde i pregiudizi dei suoi creatori e della società da cui proviene.

3. Corpi quantificati, corpi controllati: sorveglianza di genere e biopotere

Se il primo livello di rischio riguarda la discriminazione algoritmica, il secondo, più profondo, tocca la stessa materialità dei corpi e la loro soggezione a nuove forme di biopotere. Il concetto, mutuato da Michel Foucault, describe il potere sulla vita: un potere che non si limita a reprimere, ma che amministra, gestisce, ottimizza e controlla le popolazioni attraverso meccanismi di sorveglianza e regolazione. Con la sua capacità di monitorare, analizzare e classificare i corpi a livello di massa, l'IA rappresenta uno dei principali strumenti di biopotere del XXI secolo.

3.1. La sorveglianza intima: app. per il ciclo e health-tech

Un campo esemplare è quello delle tecnologie per la salute femminile. Esistono centinaia di app. per il tracciamento del ciclo mestruale, della fertilità, della menopausa. Presentate come strumenti di empowerment che restituiscono alle donne il controllo sui propri dati corporei, in realtà questi software nascondono un'altra faccia: quella della sorveglianza intima. I dati inseriti – sull'umore, sul desiderio sessuale, sulla regolarità del ciclo, sull'assunzione di contraccettivi – costituiscono un quadro biometrico e comportamentale di straordinaria delicatezza.

Il rischio è duplice. Il primo è quello della commercializzazione e della violazione della privacy. A chi sono venduti questi dati? Come vengono utilizzati? Un'indagine del Wall Street Journal ha rivelato come l'app. Flo - tra i software più popolari del momento – condividesse i dati delle utenti con Facebook e altre compagnie, senza un consenso



veramente informato (Schechner e Secada, 2019). In un'epoca in cui l'accesso all'aborto è sempre più controverso in molti paesi, inclusi gli Stati Uniti, questi dataset diventano potenziali armi di persecuzione politica. Immaginiamo un procuratore in uno stato che vieta l'aborto che richiama a un'azienda i dati delle utenti che hanno segnalato un ritardo nel ciclo: la sorveglianza digitale si trasforma direttamente in controllo sociale sui corpi delle donne.

Il secondo rischio è quello della medicalizzazione algoritmica. Questi software spesso forniscono previsioni e consigli. Ma su quali basi? Gli algoritmi che predicono l'ovulazione o i disturbi come l'endometriosi sono spesso "scatole nere" non validate dalla comunità scientifica. Possono creare ansie ingiustificate oppure, al contrario, dare un falso senso di sicurezza. Il corpo femminile, già storicamente oggetto di controllo medico e normativo, viene qui ri-mediato attraverso un algoritmo che ne pretende una conoscenza totale e predittiva, con effetti potenzialmente erosivi dell'agency e dell'intelligenza corporea delle donne stesse.

3.2. Sorveglianza lavorativa e produttività di genere

Un altro volto del biopotere algoritmico si manifesta nel mondo del lavoro, in particolare nei settori a forte femminilizzazione come il lavoro di cura, il customer service e alcuni segmenti della logistica. In questi contesti, l'IA si fa strumento di management algoritmico: sistemi che monitorano ogni movimento, ogni pausa, ogni minuto di inattività dei lavoratori, assegnando punteggi di performance e automatizzando i licenziamenti.

Per le cassiere di un supermercato, un algoritmo può misurare la velocità di scansione dei prodotti; per le operaie in un magazzino Amazon, i sensori tracciano i tempi di prelievo e suggeriscono (o impongono) percorsi ottimizzati; per le impiegate in smart working, software di *productivity monitoring* catturano screenshot dello schermo e misurano l'attività della tastiera. Questi sistemi, spacciati per oggettivi, ignorano completamente le specificità di genere. Non tengono conto del fatto che una donna



potrebbe essere interrotta più spesso da colleghi maschi (un fenomeno noto come *mansplaining* o interruzione asimmetrica) oppure che potrebbe doversi occupare, durante il lavoro da casa, di compiti di cura familiare.

Il risultato è una forma di disciplina di genere digitale. Il modello di "lavoratore ideale" incorporato in questi algoritmi è un modello maschile, astratto, libero da obblighi di cura, che può dedicare la sua intera esistenza alla produttività aziendale. Le donne, il cui tempo è storicamente più frammentato e "poroso", vengono sistematicamente penalizzate da questi sistemi, che naturalizzano un'idea di produttività profondamente sessista.

3.3. Sesso-robot e l'oggettivazione algoritmica

Un'ulteriore, inquietante frontiera del controllo dei corpi è rappresentata dai sex-robot che incorporano soluzioni di IA. Presentati dalla narrativa commerciale come compagni perfetti, sempre disponibili e privi di complicazioni, queste tecnologie sono molto spesso progettate con corpi iper-sessualizzati e femminili, e programmati con personalità sottomesse e accondiscendenti (Richardson, 2015).

Non si tratta di una semplice evoluzione tecnologica dei sex-toy, sembra piuttosto interpretabile come la materializzazione di un fantasma patriarcale: la donna-oggetto, totalmente controllabile, la cui unica funzione è soddisfare i desideri (maschili) senza porre limiti o esprimere una propria volontà. In questo caso, l'IA non si limita a riflettere uno stereotipo, ma lo istituzionalizza in una forma tangibile e interattiva. Rischia di normalizzare relazioni asimmetriche e prive di consenso, insegnando implicitamente che i corpi femminili sono territori da possedere e programmare, non partner con cui relazionarsi in modo reciproco. È la forma più estrema di biopotere algoritmico: il potere non solo di sorvegliare il corpo, ma di crearlo ex-novo secondo i propri desideri distorti.



4. L'angelo del focolare digitale: stereotipi di genere nei modelli linguistici e generativi

L'esplosione dei Grandi Modelli Linguistici (LLM) come GPT-4 e dei modelli generativi per immagini come DALL-E e Midjourney ha portato il problema degli stereotipi di genere su un palcoscenico globale e accessibile a tutti. Questi modelli, in grado di generare testi, immagini e conversazioni di sorprendente fluidità, non sono creatori *ex nihilo*. Sono specchi, potentissimi e distorti, della cultura da cui sono nati.

4.1. Linguaggio e worldview: il pregiudizio incarnato nella parola

Come accennato, gli LLM apprendono da enormi corpora testuali del web. In questi testi, gli stereotipi di genere non sono l'eccezione, ma la regola. Il risultato è che i modelli li interiorizzano e li riproducono in modo sistematico. Gli studi sui word embeddings (rappresentazioni matematiche delle parole) mostrano come i modelli associno automaticamente "uomo" con "programmatore", "dottore", "capo", e "donna" con "infermiera", "maestra", "casalinga" (Bolukbasi et al., 2016).

Questo non è un errore di calcolo; è un riflesso accurato delle associazioni statistiche presenti nella lingua. La lingua, come ci insegnano i linguisti, non è solo un mezzo di comunicazione, ma un sistema che struttura il nostro pensiero e il nostro *worldview*. Quando un modello di IA genera testo, sta generando una visione del mondo basata sulle probabilità del suo dataset. Se in quel dataset le donne sono statisticamente associate a ruoli di cura e subalterni, il "mondo" generato dall'IA sarà un mondo di matrice patriarcale.

Il rischio è la banalizzazione algoritmica degli stereotipi. Se milioni di persone iniziano a utilizzare ChatGPT per scrivere e-mail, discorsi, articoli o storie, il modello fungerà da moltiplicatore e normalizzatore di questi stereotipi. Uno scrittore in cerca di ispirazione per un personaggio femminile potrebbe ricevere dall'IA descrizioni che enfatizzano



l'aspetto fisico e l'emotività, mentre per un personaggio maschile potrebbe ottenere descrizioni focalizzate sulla competenza e l'azione. L'IA diventa così un collaboratore reazionario, che spinge inconsciamente la creatività umana verso binari già ampiamente percorsi.

4.2. L'immagine (s)regolata: i bias nei modelli generativi visivi

Ancora più evidente è il problema nei modelli di generazione di immagini. Ricerche condotte su Stable Diffusion e DALL-E hanno dimostrato che, quando si chiede al modello di generare un'immagine di una "persona" in una professione di prestigio (come "amministratore delegato", "giudice" o "scienziato"), la stragrande maggioranza delle immagini raffigura uomini, spesso bianchi. Al contrario, richieste come "assistente", "infermiere" o "persona che cucina" producono prevalentemente figure femminili (Bianchi et al., 2023).

Questo non avviene perché gli ingegneri di OpenAI o Stability AI sono sessisti, ma perché i dataset di addestramento (spesso composti da miliardi di immagini scrape-ate da internet) sono intrisi di questi squilibri. Le foto di CEO sui siti aziendali sono per lo più di uomini; le rappresentazioni stock di scienziati sono maschili; i ritratti storici sono dominati da figure maschili. Il modello impara che "autorità" e "competenza" hanno un volto maschile e, quando generiamo un'immagine, il modello sta semplicemente estraendo una rappresentazione media del concetto che gli abbiamo chiesto, riproponendo le diseguaglianze che definiscono la nostra cultura visiva.

4.3. Gli assistenti vocali e la servitù femminilizzata

Forse l'esempio più quotidiano e normalizzato di stereotipo di genere riprodotto dall'IA ci viene restituito dall'attività degli assistenti vocali: Siri di Apple, Alexa di Amazon, Google Assistant. La stragrande maggioranza di questi assistenti ha una voce femminile



di default, un tono giovane, amichevole e, soprattutto, accondiscendente.

Studi etnografici hanno evidenziato come gli utenti tendano a dare ordini bruschi e talvolta verbalmente abusivi a questi assistenti ("Siri, sei stupida!"). La risposta dell'assistente è sempre educata, servizievole, non reagisce mai. Questa dinamica riproduce e rinforza un archetipo profondamente radicato: quello della donna come segretaria, assistente, serva – una figura il cui ruolo è eseguire ordini senza discutere, sopportare gli sfoghi altrui e essere sempre disponibile (West, Kraut, e Chew, 2019).

Progettando sistemi che incarnano questo stereotipo, le aziende tech stanno ingegnerizzando la cortesia di genere. Stanno insegnando a un'intera generazione, inclusi i bambini, che è appropriato parlare in un certo modo a un'entità con una voce femminile. Stanno normalizzando una relazione di subalternità. Il fatto che alcune aziende abbiano introdotto voci maschili come opzione non risolve il problema di fondo: la scelta iniziale e predefinita rimane quella che associa l'assistenza, la disponibilità e la non-minaccia al femminile.

5. La cattedrale e il bazaar: asimmetrie di potere nell'industria tech

Per comprendere le ragioni della larga diffusione di questi bias, non basta guardare ai dati e agli algoritmi. Bisogna guardare anche alla composizione socio-economica di chi lavora nell'industria tech giocando un ruolo decisivo nella creazione delle moderne intelligenze artificiali. L'industria tecnologica globale è caratterizzata da una profonda omogeneità di classe, culturale e di genere, che si traduce in una cecità prospettica nella progettazione delle tecnologie.

5.1. Il problema della pipeline: dalle aule all'open space

Secondo il World Economic Forum (2023) e rapporti di aziende come Google e Microsoft, le donne rappresentano solo circa il 25-30% della forza lavoro nel settore



tech, una percentuale che crolla drasticamente quando si considerano i ruoli tecnici specializzati e di leadership, specialmente nel campo dell'IA. Squilibrio questo che ha origine nei percorsi scolastici e universitari delle popolazioni maschile e femminile.

Le cosiddette discipline STEM (Scienza, Tecnologia, Ingegneria e Matematica) sono storicamente dominati dagli uomini, un fenomeno alimentato da una complessa rete di stereotipi culturali, bias impliciti degli insegnanti e dalla mancanza di modelli di ruolo femminili. Le ragazze vengono spesso scoraggiate, consciamente o inconsciamente, dall'intraprendere percorsi tecnico-scientifici, percependoli come "maschili" o "poco creativi".

Il risultato è un deficit di pipeline: un numero sproporzionatamente basso di donne che decide di entrare nel settore. Questo ha un impatto diretto e drammatico sulla progettazione dell'IA. Se un team di ingegneri è composto principalmente da giovani uomini, con background socio-economico simile, è statisticamente probabile che quel team non consideri una serie di esigenze, prospettive e rischi che sono invece evidenti per altri gruppi demografici. Non si tratta di malafede, ma di un punto cieco collettivo.

5.2. La cultura "brogrammer" e l'esclusione informale

Oltre al problema quantitativo della pipeline, esiste un problema qualitativo di cultura aziendale. In molte aziende tech, specialmente nelle startup della Silicon Valley, si è affermata una cultura informale, iper-competitiva e spesso maschilista, soprannominata *brogrammer culture* (un neologismo che unisce "bro", amico, a "programmer"). È una cultura fatta di after-work in pub affollati, discussioni tecniche basate su un'aggressiva performatività intellettuale e dinamiche di gruppo che possono risultare ostili o alienanti per le donne e le minoranze di genere.

Questa cultura non è solo spiacevole; è un potente meccanismo di esclusione. Le opportunità di networking, le promozioni informali, le assegnazioni di progetti interessanti spesso avvengono in questi spazi informali, da cui molte donne si auto-escludono o vengono escluse. Il fenomeno del *mansplaining* e quello del *hepeating*



(quando un uomo ripete un'idea espressa da una donna ricevendo poi il credito per essa) sono sintomi di questo ambiente, che erode la fiducia e limita la crescita professionale delle donne (Saxenian, 1994).

Il risultato è che anche le poche donne che entrano nel settore hanno una probabilità più alta di abbandonarlo, creando un circolo vizioso. Questa emorragia di talenti femminili priva l'industria non solo di diversità, ma anche di competenze critiche e di quella pluralità di visioni necessaria per costruire tecnologie al servizio della collettività.

5.3. Il mito del genio solitario e l'economia dell'attenzione

Questa asimmetria di potere è rinforzata da una narrazione tossica: il mito del genio solitario, tipicamente maschile, che nel suo garage crea il futuro. Da Steve Jobs a Elon Musk, l'immaginario collettivo celebra la figura del (maschio) visionario e distruttivo. Questa narrazione oscura la natura profondamente collettiva e sociale del lavoro tecnico. Ogni innovazione poggia sulle spalle di una comunità di ricercatori, ingegneri, ma anche di lavoratori della catena di montaggio e di curatori di dati, spesso donne e persone del Sud Globale.

Questa narrazione, unita alla cultura programmer, contribuisce a creare un'economia dell'attenzione profondamente sbilanciata. Le conferenze tech, i panel, i premi sono dominati da uomini. Le voci maschili sono percepite come più autorevoli, le loro idee come più innovative. Questo bias di conferma si traduce in decisioni di prodotto e di investimento che perpetuano lo status quo. I venture capitalist, anch'essi prevalentemente uomini, tendono a finanziare startup fondate da uomini, con business plan che rispecchiano i loro interessi e le loro esperienze di vita. È un ecosistema chiuso che si auto-alimenta, del quale l'IA rappresenta la tecnologia di frontiera e il prodotto più puro di questo ecosistema.



6. Il Leviatano algoritmico: sorveglianza di stato e controllo sociale

Quando le tecnologie di IA sviluppate in un'industria omogenea e sessista vengono adottate dagli Stati, il rischio di genere assume una dimensione politica e autoritaria. I governi di tutto il mondo stanno implementando sistemi di IA per la sicurezza pubblica, l'allocazione dei benefici sociali e il controllo delle frontiere. In questi contesti, l'IA può diventare un potente strumento di sorveglianza differenziata e di discriminazione istituzionalizzata contro le donne e le persone migranti.

6.1. Il welfare automatizzato e la punizione delle donne povere

Il caso del "Robodebt" in Australia è diventato tristemente celebre come applicazioni dell'IA nel welfare. Si tratta di un algoritmo utilizzato per recuperare presunti debiti di welfare che calcolava erroneamente le entrate dei cittadini, causando angoscia e danni finanziari a migliaia di persone, anche con un impatto differenziato lungo le classiche direttrici della disegualianza, a partire proprio dall'appartenenza di genere.

Le donne, soprattutto le madri single, sono sovra-rappresentate tra i beneficiari di sussidi sociali e sono anche le più colpite dagli errori e dalle durezza dei sistemi automatizzati. In molti paesi, i modelli di rischio vengono utilizzati per segnalare famiglie "a rischio" per l'intervento dei servizi sociali. Questi modelli possono penalizzare le donne che mostrano pattern di comportamento giudicati "sospetti" dall'algoritmo (ad esempio, orari di lavoro irregolari, frequenti cambi di residenza), ignorando le complesse realtà della povertà e della violenza domestica. La studiosa Virginia Eubanks (2018) parla di "digital poorhouse" (il poverificio digitale): un'infrastruttura automatizzata che punisce i poveri, e tra questi, in modo particolare, le donne, sottoponendole a un controllo e una sorveglianza senza precedenti.



6.2. Sicurezza pubblica e il pregiudizio nella predizione della criminalità

I cosiddetti sistemi di polizia predittiva mappano i dati storici sui crimini per prevedere dove è più probabile che avvengano reati futuri e chi potrebbe esserne protagonista. Questi sistemi sono notoriamente viziati da un feedback loop razzista e classista: se gli eventi problematici si concentrano in quartieri a maggioranza nera e/o povera, l'algoritmo consiglierà di inviare più poliziotti lì, che si concentreranno sulle persone afferenti a gruppi marginalizzati, sostituendo il criterio di causalità con quello di correlazione nelle operazioni di contrasto alla criminalità.

Anche qui, esiste una dimensione di genere. La sorveglianza intensiva di certi quartieri colpisce le donne in modi specifici. Può rendere più pericoloso per loro, ad esempio, svolgere attività come il sex work, aumentando il rischio di arresto e violenza. Inoltre, questi sistemi sono spesso ciechi rispetto a crimini di genere come la violenza domestica, che spesso avviene in spazi privati e sono storicamente sotto-segnalati. L'algoritmo, concentrandosi sui crimini di strada, riproduce una gerarchia di visibilità che marginalizza ulteriormente le vittime di violenza di genere.

6.3. Controllo delle frontiere e corpi femminili iper-sorvegliati

Nel campo del controllo delle frontiere, l'IA viene utilizzata per analizzare i dati biometrici, valutare la "credibilità" dei richiedenti asilo e persino prevedere i flussi migratori. Anche in questo caso, le donne e i bambini sono soggetti a rischi specifici. I sistemi di riconoscimento facciale possono avere difficoltà con le donne velate, portando a ulteriori controlli e umiliazioni. I modelli per valutare le richieste di asilo potrebbero non dare il giusto peso a forme di persecuzione specifiche di genere, come le mutilazioni genitali femminili o la violenza sessuale sistematica.

La studiosa Louise Amoore (2020) ha teorizzato l'"etica algoritmica" dei confini, in cui



decisioni che hanno un impatto profondo sulla vita delle persone vengono delegate a modelli che calcolano probabilità e rischi. In questo calcolo, la specificità del corpo femminile migrante, con le sue vulnerabilità e le sue esperienze uniche, rischia di essere cancellata, ridotta a un dato anomalo in un modello progettato per una norma maschile e astratta (Amoore, 2020).

7. Oltre la critica: verso un'IA femminista e intersezionale

Di fronte a questo quadro apparentemente apocalittico, la tentazione del rifiuto luddista è forte. Ciò nonostante, l'alternativa non può essere tra adozione acritica e rifiuto totale. Esiste una terza via: reclamare l'IA come terreno di lotta politica e progettare attivamente un futuro diverso. Questo progetto passa attraverso l'elaborazione di un'etica femminista dell'IA, che sia anche radicalmente intersezionale.

7.1. I principi di un'etica femminista per l'intelligenza artificiale

Un'etica femminista per l'IA non si accontenta di aggiungere "diversità" a un sistema fondamentalmente ingiusto, mettendo invece in discussione i presupposti stessi del sistema a partire da un complesso di principi fondamentali.

Il primo principio è quello della "conoscenza situata" (Haraway, 1988), che si contrappone al mito della vista da nessun luogo (the view from nowhere) dell'oggettività scientifica. Ogni conoscenza, incluso quella prodotta da un algoritmo, è parziale, prodotta da un punto di vista specifico. Per essere etica, l'IA dovrebbe quindi essere esplicitamente situata, dichiarare i suoi limiti, i suoi punti di vista, i dati su cui è stata addestrata e cercare al contempo di superare i pericoli associati a quel particolare situarsi.

Il secondo principio concerne l'"etica della cura", sviluppata da pensatrici come Carol Gilligan (1982) e Joan Tronto (1993), mette al centro non l'individuo astratto e



autonomo, ma le reti di relazioni di interdipendenza che costituiscono la vita. Per essere etica, l'IA non dovrebbe essere progettata per ottimizzare soltanto il profitto o l'efficienza, ma il benessere relazionale, la giustizia sociale e la sostenibilità ecologica.

Il terzo principio, infine, riguarda l' "accountability". Per essere etica, l'IA dovrebbe permettere di capire come è stata presa una decisione, di contestarla e di attribuire chiaramente le responsabilità derivanti da quel processo decisionale. Questo richiede trasparenza non solo dell'algoritmo, dei processi decisionali aziendali e anche delle catene di fornitura.

7.2. L'imperativo intersezionale

Il femminismo bianco e liberale ha storicamente fallito nel rappresentare le esperienze delle donne nere, delle donne del Sud Globale, delle donne disabili e delle persone queer e non-binary. Kimberlé Crenshaw (1989) ha coniato il termine intersezionalità per descrivere come le diverse forme di oppressione (razzismo, sessismo, classismo, abilismo) si intrecciano e si moltiplicano, creando esperienze uniche di discriminazione. Per essere intrinsecamente giusta, l'IA dovrebbe essere progettata attraverso un approccio intersezionale, attento a considerare le diverse direttrici della disuguaglianza e le potenziali combinazioni.

La giustizia algoritmica richiede di considerare le identità multiple e sovrapposte degli individui, coinvolgendo nel processo di progettazione donne di diverse etnie, classi sociali, abilità e orientamenti sessuali, e costruendo dataset rappresentativi della ricchezza e della complessità dell'esperienza umana.

7.3. Pratiche concrete: participatory design e data feminism

La possibilità di tradurre queste considerazioni in pratiche concrete può passare attraverso due approcci promettenti, distinti, ma complementari.



Il *participatory Design* (Co-Design) consiste nel coinvolgimento attivo in ogni fase di progettazione di una particolare tecnologia di tutte quelle soggettività che ne saranno colpite (ad esempio, le lavoratrici di piattaforma, le donne che usano app per la salute, le richiedenti asilo): dalla definizione del problema, alla raccolta dei dati, al testing del prototipo. La partecipazione si configura, quindi, come pratica per la rottura di quelle asimmetrie di potere che separano il progettista e l'utente, garantendo che la tecnologia risponda a bisogni reali e non imposti.

Il secondo approccio è quello del cosiddetto "data Feminism" (D'Ignazio e Klein, 2020). Il femminismo dei dati punta a sfidare il potere e le sue rappresentazioni stereotipate, superando il binarismo di genere, rendendo visibile il lavoro di cura (spesso invisibile nei dataset) e socializzando la collettività all'idea che i dati non siano mai "grezzi" ma sempre "cotti" da interpretazioni e valori (D'Ignazio e Klein, 2020).

8. Progettare il futuro: governance, educazione e sfide aperte

L'agenda per un'IA di genere giusta è ambiziosa e richiede un'azione concertata a tutti i livelli: dalla regolamentazione internazionale all'educazione delle nuove generazioni, fino alla responsabilità individuale degli sviluppatori.

8.1. Regolamentare il Leviatano: GDPR, AI Act e oltre

La governance dell'IA sta lentamente emergendo. Il Regolamento Generale sulla Protezione dei Dati (GDPR) dell'Unione Europea ha stabilito principi importanti come la *privacy by design* e il diritto alla spiegazione per decisioni automatizzate. Un passo ulteriore è il AI Act europeo, il primo tentativo comprensivo di regolamentare l'IA in base al rischio, che vieta alcune pratiche considerate inaccettabili (come il social scoring dei governi) e impone obblighi di trasparenza e robustezza per i sistemi che presentano un grado di rischio elevato.



Tuttavia, queste regolamentazioni devono essere costantemente aggiornate e rafforzate per affrontare specificamente i rischi di genere. Si dovrebbe, ad esempio: richiedere un audit dell'impatto di genere per i sistemi di IA ad alto rischio; imporre la diversità dei dataset e test di performance intersezionali prima dell'immissione sul mercato; stabilire garanzie processuali forti per i cittadini colpiti da decisioni algoritmiche discriminatorie, portando l'onere della prova a gravare sul titolare del sistema piuttosto che sulla vittima.

8.2. Educare cittadini critici e sviluppatori etici

La regolamentazione da sola non basta. È fondamentale un investimento massiccio nell'educazione. Da un lato, un'educazione alla cittadinanza digitale critica che insegni a tutti, fin dalla scuola, a comprendere i meccanismi di base dell'IA, a riconoscere i bias e a difendere i propri diritti digitali. Dall'altro, una rieducazione dei curricula STEM per ingegneri e data scientist.

L'etica non può essere un corso opzionale, dovrebbe invece essere il fondamento di ogni formazione tecnologica. Gli sviluppatori del futuro dovrebbero essere formati non solo a scrivere codice efficiente, ma a porsi domande come: "Chi è escluso dal mio design?", "Quali valori sono incorporati in questo algoritmo?", "Quali sono le conseguenze a lungo termine di questa tecnologia?".

8.3. Le sfide aperte: dall'agnosticismo tecnico all'azione affermativa

Il percorso è irto di ostacoli. Il principale è la resistenza culturale dell'industria tech, che spesso si nasconde dietro un agnosticismo tecnico ("la tecnologia è solo uno strumento, siamo neutrali") per evitare di assumersi responsabilità. Superare questa resistenza richiederà una pressione costante da parte della società civile, dei giornalisti, degli accademici e degli investitori consapevoli.



In alcuni casi, potrebbe essere necessario un dibattito serio su forme di azione affermativa algoritmica. Così come esistono quote per garantire la rappresentanza di genere in politica, potremmo aver bisogno di meccanismi che correggano attivamente gli squilibri nei risultati algoritmici, almeno in transizione, per compensare secoli di discriminazione strutturale. È una proposta controversa, che solleva questioni sulla "discriminazione inversa", ma è una discussione che non possiamo evitare se vogliamo seriamente colmare il divario di genere.

Infine, c'è la sfida della complessità tecnica. Rendere gli algoritmi veramente trasparenti e accountable è estremamente difficile, specialmente per sistemi di deep learning con milioni di parametri. La ricerca in campi come l'AI explainability (XAI) è cruciale, ma non possiamo aspettare che risolva tutti i problemi prima di agire. A volte, la scelta più etica potrebbe essere quella di non automatizzare affatto un determinato processo decisionale, soprattutto quando riguarda vite umane e diritti fondamentali.

Conclusione: il futuro è una scelta, non un output algoritmico

L'intelligenza artificiale non va intesa come un destino ineluttabile. È un insieme di tecnologie progettate da esseri umani, in istituzioni umane, per scopi umani.

I rischi analizzati in questa sede – la discriminazione algoritmica, la sorveglianza di genere, la cristallizzazione degli stereotipi, l'erosione dell'agency – non sono bug tecnici, ma il sintomo di un potere mal distribuito e di una cecità culturale.

La triade genere, potere e rischio nell'era dell'IA ci impone di fare una scelta. Possiamo accettare passivamente che le tecnologie più potenti del nostro tempo siano plasmate da una piccola e omogenea élite, rischiando di codificare per l'eternità digitale le ingiustizie del passato. Oppure possiamo reclamare collettivamente il futuro. Possiamo insistere per un'IA che non sia solo "etica" in un documento formale, ma che sia femminista nei suoi principi, intersezionale nella sua progettazione, democratica nella sua governance e umana nei suoi obiettivi finali.

Il lavoro da fare è ampio e coinvolge diversi fronti: smantellare la cultura programmer,



diversificare i team, ripulire i dataset, aprire le scatole nere, regolamentare con coraggio, educare con visione. Si tratta di un lavoro politico e culturale prima che tecnico. È il lavoro di costruire un mondo in cui la macchina non serva a replicare il potere di pochi, ma a distribuire l'agency e il benessere tra tutti i generi. Il futuro non è un output da calcolare; è una scelta da fare insieme.



Riferimenti bibliografici

1. Amoores, L. (2020). *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*. Duke University Press. ISBN: 9781478008420.
2. Bender, E. M., Gebru, T., McMillan-Major, A., and Shmitchell, S. (2021). *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. <https://doi.org/10.1145/3442188.3445922>.
3. Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity Press. ISBN: 9781509526390.
4. Bianchi, F., Kalluri, P., Durmus, E., Ladhak, F., Cheng, M., Nozza, D., Hashimoto, T., Jurafsky, D., Zou, J., and Caliskan, A. (2023). *Easily Accessible Text-to-Image Generation Amplifies Demographic Stereotypes at Large Scale*. Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, 1493–1504. <https://doi.org/10.1145/3593013.3594095>.
5. Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., and Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Advances in Neural Information Processing Systems*, 29: 4349–4357. <https://arxiv.org/abs/1607.06520>
6. Browne, S. (2015). *Dark Matters: On the Surveillance of Blackness*. Duke University Press. ISBN: 9780822359384.
7. Buolamwini, J., and Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of Machine Learning Research, 81, 1–15. <http://proceedings.mlr.press/v81/buolamwini18a.html>.
8. Butler, J. (1990). *Gender Trouble: Feminism and the Subversion of Identity*. Routledge. ISBN: 9780415389556.
9. Cheryan, S., Plaut, V. C., Davies, P. G., & Steele, C. M. (2009). Ambient belonging: How stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology*, 97(6), 1045–1060. <https://doi.org/10.1037/a0016239>.
10. Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press. ISBN: 9780300209570.
11. Crenshaw, K. (1989). Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. *University of Chicago Legal Forum*, 1989(1), 139–167. <https://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8>.
12. Dastin, J. (2018). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. <https://www.reuters.com/article/idUSKCN1MK0AG/>.
13. D'Ignazio, C., and Klein, L. F. (2020). *Data Feminism*. The MIT Press. ISBN: 9780262044004.
14. Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish*



- the Poor*. St. Martin's Press. ISBN: 9781250074317.
15. European Commission. (2021). Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>.
 16. European Parliament and Council. (2016). Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). <https://eur-lex.europa.eu/eli/reg/2016/679/oj>.
 17. Foucault, M. (1976). *La volontà di sapere. Storia della sessualità 1*. Feltrinelli. ISBN: 9788807880730.
 18. Gebru, T. (2020). *Race and Gender*. In *The Oxford Handbook of Ethics of AI* (pp. 253-274). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190067397.013.13>.
 19. Gilligan, C. (1982). *In a Different Voice: Psychological Theory and Women's Development*. Harvard University Press. ISBN: 9780674445444.
 20. Google LLC. (2023). 2023 Diversity Annual Report. <https://shorturl.at/PQZ7j>.
 21. Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, 14(3), 575–599. <https://doi.org/10.2307/3178066>.
 22. Hooks, B. (2000). *Feminism Is for Everybody: Passionate Politics*. South End Press. ISBN: 9780896086289.
 23. Jobin, A., Ienca, M., and Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. <https://doi.org/10.1038/s42256-019-0088-2>.
 24. Lupton, D. (2016). *The Quantified Self: A Sociology of Self-Tracking*. Polity Press. ISBN: 9780745667852.
 25. Microsoft (2023). 2023 Diversity & Inclusion Report. <https://www.microsoft.com/en-us/diversity/inside-microsoft/annual-report>.
 26. Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press. ISBN: 9781479837243.
 27. O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishing Group. ISBN: 9780553418811.
 28. Richardson, K. (2015). The Asymmetrical 'Relationship': Parallels Between Prostitution and the Development of Sex Robots. *ACM SIGCAS Computers & Society*, 45(3), 290–293. <https://doi.org/10.1145/2874239.2874281>.
 29. Sap, M., Card, D., Gabriel, S., Choi, Y., and Smith, N. A. (2019). *The Risk of Racial Bias in Hate Speech Detection*. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 1668–1678. <https://doi.org/10.18653/v1/P19-1163>.
 30. Saxenian, A. (1994). *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Harvard University Press. ISBN: 9780674753402.



31. Schechner, S., and Secada, M. (2019, 22 febbraio). *You Give Apps Sensitive Personal Information. Then They Tell Facebook*. The Wall Street Journal. <https://shorturl.at/Ylr6e>.
32. Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., and Vertesi, J. (2019). *Fairness and Abstraction in Sociotechnical Systems*. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), 1–36. <https://doi.org/10.1145/3351095>.
33. Stark, L., and Levy, K. (2018). The Surveillant Society of the Digitally Mediated Public Sphere. *Surveillance & Society*, 16(2), 242-248. <https://doi.org/10.24908/ss.v16i2.12944>.
34. Tronto, J. C. (1993). *Moral Boundaries: A Political Argument for an Ethic of Care*. Routledge. ISBN: 9780415906425.
35. Vallor, S. (2016). *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford University Press. ISBN: 9780190498511.
36. Wachter, S., Mittelstadt, B., and Floridi, L. (2017). Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. *International Data Privacy Law*, 7(2), 76–99. <https://doi.org/10.1093/idpl/ix005>.
37. West, M., Kraut, R., and Chew, H. E. (2019). *I'd blush if I could: closing gender divides in digital skills through education*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1>.
38. World Economic Forum. (2023). *Global Gender Gap Report 2023*. <https://www.weforum.org/publications/global-gender-gap-report-2023/>.
39. Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs. ISBN: 9781610395694.