

Call for papers: "Intelligenza  
Artificiale: prospettive bioetiche,  
bio giuridiche e sociali"

# Ingiustizia epistemica, intelligenze artificiali e sfide morali

## *Epistemic Injustice, Artificial Intelligence, and Moral Threats*

SIMONA TIRIBELLI  
simona.tiribelli@unimc.it

AFFILIAZIONE  
Università degli Studi di Macerata, Macerata, Italia  
Institute for Technology & Global Health, Boston, US

## SOMMARIO

Nelle società dell'informazione odierne, i sistemi di intelligenza artificiale (IA) appaiono detenere un vantaggio epistemico sugli individui, poiché capaci di processare, dare senso e inferire informazioni preziose dalla realtà datificata in cui viviamo, noi inclusi. Tale vantaggio ha funto da sprone al loro uso pervasivo al punto che oggi co-partecipano e rimodellano gran parte dei nostri compiti epistemici e delle nostre decisioni. Se tale ruolo e impatto sono riconosciuti, se tali sistemi siano vantaggiosi o dannosi per gli individui dal punto di vista epistemico è materia di indagine. Questo articolo si propone di chiarire i benefici e, in modo particolare, i danni producibili dalle tecnologie algoritmiche sugli individui quali soggetti epistemici, traendo spunti dalle teorie sull'ingiustizia epistemica elaborate nell'ambito della filosofia morale e dall'epistemologia sociale. In tal modo, l'articolo mira anche a chiarire le sfide morali intrinseche a tali danni e, infine, sottolineare il ruolo della filosofia morale nell'etica dell'IA per affrontarli.

## PAROLE CHIAVE

Ingiustizia epistemica  
Intelligenza artificiale  
Etica dell'intelligenza artificiale  
Agency epistemica  
Danni epistemici

## ABSTRACT

*Artificial intelligence (AI) systems sound to hold an epistemic advantage on individuals in nowadays mature information societies, insofar as they are capable of navigating, making sense, and inferring valuable knowledge from the datified reality we live in, ourselves included. Such advantage spurred their pervasive use so that today AI systems co-participate and reshape the majority of epistemic tasks and decisions we form and take. While such a role and impact are recognized, whether they benefit or harm individuals epistemically ought to be investigated. This paper aims to elucidate the epistemic challenges posed by algorithmic technology to individuals in their epistemic standing, by drawing on theories on epistemic injustice developed in moral philosophy and social epistemology. By doing so, the paper aims also to clarify the moral threats entrenched with the AI-based epistemic harms outlined and to finally point out the role of moral philosophy for AI ethics to tackle them.*

## KEYWORDS

*Epistemic Injustice  
Artificial Intelligence  
AI Ethics  
Epistemic Agency  
Epistemic Harms*

DOI: 10.53267/20240101



Call for papers: "Intelligenza Artificiale: prospettive bioetiche, biogiuridiche e sociali"

## 1. INTRODUZIONE

Sistemi di intelligenza artificiale (IA) come algoritmi di *machine learning* e *deep learning* governano oggi in modo pervasivo gran parte delle tecnologie digitali con cui facciamo esperienza e conosciamo il mondo, diamo senso alle nostre esperienze, apprendiamo e ci esprimiamo, comunichiamo con gli altri e persino con noi stessi. In altre parole, le tecnologie basate su tali sistemi co-partecipano in vari ruoli a innumerevoli nostre attività di natura *epistemica*. Secondo alcuni, questi sistemi costituiscono un valido *supporto epistemico*, date le nostre comprovate limitate capacità cognitive, soprattutto negli ambienti delle società mature dell'informazione, caratterizzati da sovraccarico informativo e scarse risorse temporali<sup>1</sup>. Altri, invece, li descrivono come strumenti di *oppressione epistemica* e ingiustizia sociale, essendone stata dimostrata la tendenza a incorporare pregiudizi iniqui storicamente radicati (quali i *bias* etnici e di genere), difficili da scovare a causa della relativa *opacità intrinseca*, e così di perpetuare *modelli epistemici di discriminazione sistemica*, influenzando in modo strutturale il modo in cui conosciamo, interpretiamo esperienze, formiamo credenze, comunichiamo e agiamo, perlopiù a discapito di alcuni gruppi minoritari<sup>2</sup>.

Tuttavia, se il rapporto tra sistemi di IA e ingiustizia sociale è oggi al centro di un'ampia letteratura nell'ambito dell'etica delle IA, soprattutto negli studi sull'equità algoritmica, la relazione tra IA e ingiustizia epistemica è trattata in modo parziale e frammentario.

Questo articolo intende mostrare come le lenti teoriche elaborate in materia di ingiustizia epistemica ci consentano di elaborare un quadro sistematico e chiaro di quali siano le sfide di matrice epistemica che i sistemi di IA possono inaugurare, perpetuare ed esacerbare, nonché il portato morale a queste intrinseco. A tal fine, la prima sezione mostra la rilevanza del tema a partire dal dibattito in materia; la seconda sezione introduce la tassonomia concettuale di analisi delle sfide e dei possibili danni epistemici introdotti o riprodotti dalle IA, e li discute, traendo spunti dagli studi più dirimenti sull'ingiustizia epistemica elaborati in filosofia morale e nell'epistemologia sociale; la terza sezione mostra la rilevanza morale di tali sfide e conclude sottolineando il ruolo chiave della filosofia ed epistemologia morale nell'etica dell'IA al fine di affrontarle.

## 2. INGIUSTIZIA EPISTEMICA NELL'ETICA DELL'IA

Il tema dell'ingiustizia è uno tra più discussi negli studi sull'etica dell'IA. Quivi il dibattito si è concentrato principalmente su questioni di ingiustizia sociale; questioni innescate e perlopiù connesse alla tendenza dei sistemi di IA a incorporare *bias* iniqui nei set di dati utilizzati per svilupparli (*bias di minoranza* inclusi: tendenza all'uso di set di dati mancanti di una rappresentazione adeguata di gruppi vulnerabili e minoranze) e, quindi, a riprodurre ed esacerbare nei relativi risultati (decisioni, punteggi, raccomandazioni) asimmetrie di potere, oppressione sociale e disuguaglianze inique socialmente stratificate e storicamente radicate. Rare sono le indagini filosofiche di matrice etico-applicata sul particolare tipo di ingiustizia nota, invece, come *ingiustizia epistemica* in relazione alle tecnologie in questione<sup>3</sup>. Chiariamo, in primo luogo, perché una tale indagine appare rilevante.

Nel suo celebre lavoro "Epistemic Injustice: Power and the Ethics of Knowing", Miranda Fricker definisce le ingiustizie epistemiche come quei torti o danni che minano gli individui quali *soggetti epistemici (knowers)*, cioè, nelle proprie capacità epistemiche (di conoscere, interpretare, comunicare, nonché essere ascoltati e creduti)<sup>4</sup>. L'impatto dei sistemi algoritmici sugli individui come soggetti epistemici (conoscitivi e decisionali) appare oggi evidente. Si pensi, ad esempio, al ruolo mai neutrale che esercitano quali *gatekeepers* dell'informazione, già centrale nelle piattaforme digitali quali motori di ricerca e *social networking services* e ora ancora più critico considerati i progressi nell'ambito dei *large language models* (e delle IA generative in senso ampio). Inoltre, se, come è noto, le tecnologie algoritmiche sono le sole, oggi, a esibire le capacità necessarie per processare e dare senso agli enormi flussi di dati risultanti dai processi di digitalizzazione e datificazione del reale (tutto è datificabile: incorporabile o traducibile in dati) e, da ciò, produrre informazioni, anche di carattere predittivo, di grande valore nei processi economici e politici odierni, appare altresì evidente il *vantaggio epistemico (epistemic advantage)*, o *posizione epistemica privilegiata*, di tali sistemi e di chi li possiede e ne trae profitto su scala<sup>5</sup>. Tuttavia, se questo vantaggio epistemico detenuto e prodotto dalle IA produca dei benefici o danni sulle capacità epistemiche umane (conoscitive e decisionali), per cui si possa parlare

di ingiustizie, e con quali implicazioni etiche, richiede una riflessione sistematica ben precisa.

Vari studi nell'etica digitale e dell'IA hanno descritto tali sistemi, incluse le tecniche da questi dispiagate (tecniche di personalizzazione, dalla profilazione algoritmica, filtro e classificazione ai sistemi di raccomandazione), come *abilitatori cognitivi ed epistemici*<sup>6</sup>. Ciò appare ragionevole almeno per due ragioni. In primo luogo, consentono di far fronte alle nostre ridotte capacità cognitive di processare e scegliere in contesti caratterizzati da un numero elevato di opzioni (cioè informazioni) e alle limitate risorse umane temporali ed epistemiche (assenza di tempo e informazioni rilevanti) nei contesti quotidiani, presentandoci, filtrando, a battuta di click, ciò che cerchiamo e potrebbe servirci, potenzialmente prevenendo l'utilizzo di scorciatoie cognitive o euristiche errate a cui siamo, invece, avvezzi. A tal riguardo, un numero di studi valorizza come tali sistemi permettano di rilevare noti e/o nuovi *bias* cognitivi, fallacie epistemiche e *pregiudizi iniqui* umani, difficili da identificare e rilevare come *patterns epistemic* ricorrenti, consentendoci una maggiore *vigilanza epistemica*: un'attenzione critica sui nostri processi epistemic interni ed esterni, cioè di acquisizione, processazione e condivisione della conoscenza, ad esempio, nella formazione delle credenze e relativa comunicazione e condivisione nei processi di dialogo, argomentazione e giustificazione reciproca interpersonale. In secondo luogo, tali sistemi consentono di leggere e rappresentare il reale (invisibile) su scala micro e macro e scoprire conoscenza (risoluzione di problemi inclusa) fondamentale al progresso scientifico e sociale in molti ambiti (si pensi all'ambito medico) attraverso *operazioni epistemiche complesse*, dall'analisi dati all'individuazione di modelli, passando per processi di inferenza e scoperta di evidenza, indispensabili per le capacità cognitive ed epistemiche umane naturali.

Altri studiosi sostengono che tali sistemi siano più di meri abilitatori cognitivi; sono tecnologie *intrinsecamente epistemiche*: sistemi progettati per funzionare in ambienti *epistemic*, come gli ambienti informazionali, utilizzati su *contenuti epistemic*, come dati e informazioni, e implementati per operare attraverso *operazioni epistemiche* come analisi, inferenze e predizioni, senza necessariamente migliorare le nostre capacità epistemiche (possono generare conoscenza inesatta in modo opaco)<sup>7</sup>.

In sintesi, le IA hanno capacità fondamentalmente epistemiche, in ultima istanza, analizzano dati, e generano benefici e danni che sono, quindi, *in primis*, di natura epistemica. È all'analisi di questi danni che dedichiamo la prossima sezione. Utilizzando le lenti teoriche elaborate nell'ambito degli studi sull'ingiustizia epistemica, vagliamo se le tecnologie algoritmiche non solo possano promuovere le nostre capacità epistemiche, come emerge *prima facie*, ma pure danneggiare la nostra agency epistemica, generando danni epistemic ingiusti.

Capire la natura dei danni riprodotti e generati dai sistemi di IA è di fondamentale importanza: ci consente non solo di chiarire perché (giustificazione morale) i sistemi che li producono dovrebbero essere regolati; ci consente soprattutto di elaborare in modo realmente efficace *come* farlo e, nello specifico, attraverso quali strumenti teorici e pratici, per prevenire o mitigare i rischi in questione.

### 3. IA E DANNI EPISTEMICI

Nei casi di ingiustizia epistemica è sempre coinvolto un certo tipo di danno epistemico, cioè un danno arrecato a qualcuno quale agente epistemico. Fricker, il cui merito è *in primis* quello di aver distinto nel suo lavoro questa particolare ingiustizia, definisce due tipologie di ingiustizia epistemica: *l'ingiustizia testimoniale* e *l'ingiustizia ermeneutica*. *L'ingiustizia testimoniale* si verifica quando a un parlante è dato un inferiore livello di credibilità a causa di un pregiudizio identitario; si pensi ai molti esempi di casi in cui la parola di un testimone di un crimine non è presa sul serio a causa di un *bias* etnico. *L'ingiustizia ermeneutica* si verifica, invece, «quando una lacuna nelle risorse interpretative collettive mette qualcuno in una posizione di svantaggio ingiusto quando si tratta di dare un senso alle proprie esperienze sociali»<sup>8</sup>. Uno degli esempi più noti di ingiustizia ermeneutica è il caso di Carmita Wood: una vittima di molestia sessuale nel 1970, incapace di comprendere e, quindi, comunicare la propria esperienza a causa dell'assenza di un concetto adeguato per interpretare e comunicare l'ingiustizia subita. Fricker distingue inoltre tra *danni epistemic primari* (intrinseci) e *secondari* (estrinseci). Il danno primario dell'ingiustizia epistemica consiste nell'essere danneggiati *qua knower* e quindi simbolicamente *qua personae*, in altre parole, quali agenti epistemic capaci di partecipare ai processi di produzione di conoscenza;

Ingiustizia  
epistemica

Call for papers:  
"Intelligenza  
Artificiale:  
prospettive  
bioetiche,  
biogiuridiche e  
sociali"

Volume 9 ■ 2024

theFuture  
ofScience  
andEthics

17

quivi, secondo Fricker, vi è una riduzione ingiusta del *soggetto* da *agente* epistemico *attivo* a *oggetto* epistemico *passivo*, da cui esclusivamente estrarre conoscenza (processo che Fricker chiama di *oggettificazione epistemica*). Non tutti concordano sul fatto che risieda nell'oggettificazione il danno primario dell'ingiustizia epistemica; altri sostengono che sia invece nell'essere visti come soggetto inaffidabile, incapace di contribuire alle pratiche epistemiche in modo *unico*, cioè a partire dalle proprie *particolari* esperienze vissute<sup>9</sup>. I danni secondari dell'ingiustizia epistemica si distinguono in a. danni secondari pratici e b. danni secondari epistemici. I primi si riferiscono ai costi pratici causati dall'ingiustizia in questione. Ad esempio, l'assenza di risarcimento per un danno subito a causa di una sentenza negativa dovuta a un pregiudizio etnico o, in parte, a causa dell'incapacità della persona di interpretare e poi comunicare l'esperienza subita. I secondi si riferiscono alla perdita di fiducia nelle proprie convinzioni e competenze epistemiche, ad esempio, sulla percezione e peso di ciò che hanno vissuto, successiva all'ingiustizia epistemica subita.

Chiarita, in breve, la tassonomia di analisi di cui ci avvaliamo, è possibile elaborare alcune considerazioni. In effetti, è possibile interpretare molti dei rischi generati e riprodotti dai sistemi di IA come sfide o danni di natura epistemica. Ad esempio, i sistemi di IA appaiono perpetuare una condizione di *ingiustizia epistemica ermeneutica by default* a causa dell'*opacità epistemica* a essi intrinseca; in altre parole: la loro complessità e seguente non intellegibilità priva gli individui degli strumenti concettuali necessari per comprendere le proprie esperienze come mediate, informate o rimodellate dagli algoritmi. In effetti, il primo limite epistemico di tali sistemi è nella relativa opacità a causa della complessità del processo probabilistico in atto. Ciò costituisce una delle preoccupazioni maggiori essendo utilizzati in misura crescente per compiti e decisioni prima esclusivamente umani in settori fondamentali quali l'educazione e la sanità. Come è noto, il 'comportamento' algoritmico, cioè il modo attraverso il quale un sistema elabora un certo *output* (il quale può assumere varie forme: punteggio, risultato, decisione, raccomandazione, a seconda del sistema considerato), processando probabilisticamente enormi quantità di dati, non solo è spesso inaccessibile per ragioni quali la proprietà intellettuale, ma – anche nei

casi di modelli *open source* – risulta incomprensibile per natura probabilistica ai suoi stessi designer; in altre parole, è una 'scatola nera'<sup>10</sup>. Questa opacità mina *by default* gli individui come agenti *epistemici* (*knowers*), cioè nella relativa possibilità (*epistemic standing*) di conoscere e comprendere il modo, le ragioni e gli aspetti (e, nella terminologia specifica, i *patterns* e le correlazioni) sulla base dei quali sono formulate certe decisioni e le relative conseguenze a cui sono soggetti. In questo senso, se da un lato tali sistemi appaiono *prima facie* agevolare le nostre scelte svolgendo alcune operazioni epistemiche quali la processazione delle informazioni, i processi algoritmici sulla base dei quali questi operano fornendo conoscenza, suggerendoci certe informazioni invece di altre, o plasmando i nostri contesti informativi, sono opachi. Tale opacità si aggrava poiché non solo riguarda i processi per ragioni di complessità probabilistica (opacità *tecnica*), ma anche i fini che li orientano. In altre parole, nonostante il sistema si comporti in modo opaco, il fine verso cui tale comportamento si dispiega (il sistema apprende probabilisticamente come raggiungerlo) è, invece, definito da alcuni (*provider*), secondo le relative agende, in modo opaco agli utenti. La conoscenza dei fini in questione mitigherebbe l'opacità tecnica in questione, ma è preclusa agli utenti, perlopiù lasciati *out of the loop* nella definizione delle agende economiche dei *provider*, a loro discapito: si tratta, dunque, di un'opacità anche *politica*. In tal senso, l'utente è privato *by design* della conoscenza sia per comprendere che cosa (e perché) sta contribuendo a plasmare le proprie esperienze sociali e credenze personali e per partecipare a tali processi, sia per riconoscere possibili fenomeni di ingiustizia a cui è soggetto – ad esempio, l'esclusione nell'esposizione a certe informazioni (beni epistemici) o a opportunità sociali ed economiche che tali informazioni incorporano (ad esempio, una posizione professionale) a causa di un'errata profilazione basata su di un pregiudizio identitario<sup>11</sup>.

Questa duplice opacità produce danni secondari epistemici e pratici rilevanti. Infatti, mina altresì gli individui come agenti *pratici*, nella possibilità di agire a riguardo di tali decisioni, *sociali* e *politici*, nella possibilità di contestarle e sovvertirle, e *morali*, nella possibilità – adeguatamente informata – di avallarne il portato valoriale intrinseco. Ciò, oggi, è di estrema rilevanza: basti pensare agli esempi di usi dei sistemi di IA per



prevedere la possibilità di recidiva di un incriminato in base a cui decidere sulla relativa libertà vigilata o per determinarne l'accesso prioritario a cure mediche. L'impossibilità di conoscere e contestare tali processi può risultare in esclusioni o privazioni pratiche sostanziali e ingiuste (ad esempio, la possibilità di un risarcimento se erroneamente valutati da un algoritmo in un caso giudiziario, si consideri il caso COMPAS, o l'accesso a una posizione professionale). Inoltre, l'incapacità di interpretare in modo adeguato e dimostrare l'ingiustizia subita può provocare danni secondari epistemici problematici, dalla perdita di fiducia rispetto alle proprie convinzioni (es. di aver ricevuto un'inesatta prevalutazione algoritmica) a un generale senso di impotenza in conoscenza e azione. Questi danni rinforzano in modo particolare quei fenomeni di marginalizzazione sociale che affliggono soprattutto le minoranze e coloro già vulnerabili, privandoli ulteriormente degli strumenti concettuali e pratici per riconoscere, interpretare e avvalorare, prima, e poi manifestare e rivendicare, le proprie particolari esperienze di ingiustizia e oppressione vissute. L'opacità duplice analizzata, infatti, crea, acuisce e cela effetti ingiusti di natura ermeneutica che, per l'appunto, tendono a colpire soprattutto gruppi storicamente marginalizzati (minoranze etniche e di genere). Tale opacità maschera una progettazione tecnologica dei sistemi in questione non inclusiva bensì storicamente escludente di gruppi socialmente marginalizzati, basata perlopiù su approcci, agende politiche ed economiche, priorità e preferenze orientate alla massimizzazione del profitto e di stampo esclusivamente occidentale, risultando nello sviluppo di sistemi che generano – come prodotto diretto – l'inevitabile rafforzamento di quei fenomeni di marginalizzazione ed esclusione sociale che ne hanno improntato le pratiche di sviluppo<sup>12</sup>.

Tali considerazioni consentono di mettere in luce ulteriori danni epistemici riconducibili a ingiustizie testimoniali. Come ampiamente mostrato da Zuboff nel suo lavoro "The Age of Surveillance Capitalism", gli algoritmi trattano gli individui (e i loro dati) come materia grezza da cui ricavare o estrarre prodotti di previsione comportamentale da vendere al miglior inserzionista<sup>13</sup>. Ciò avviene in modo particolare nei processi di profilazione individuale attraverso cui gli algoritmi inferiscono caratteristiche simili tra individui (*collaborative filtering*) al fine di assegnare loro profili (processo di etichettatura) utili al loro inse-

rimento all'interno di categorie predefinite (a cui corrispondono stime di identità e comportamenti) basate su macro-generalizzazioni. In questi processi, l'individuo non ha quasi mai un ruolo attivo, cioè non è posto nelle condizioni di partecipare in modo proattivo ai meccanismi attraverso cui è 'identificato' algoritmicamente e sulla base di cui (quali correlazioni e *patterns*) risulta esposto ad alcune informazioni e opportunità e soggetto a certe decisioni, rispetto ad altre. In questo senso, utilizzando la terminologia di Fricker, appare verificarsi un vero e proprio processo di *oggettificazione algoritmica* dell'individuo che risulta minato nella sua capacità agenziale epistemica di fornire conoscenza (agente epistemico attivo) e trattato, invece, quale – o ridotto a – *oggetto epistemico passivo* da cui ricavare conoscenza utile ai profitti di parti terze. Si tratta di un danno epistemico primario al centro di un'ingiustizia testimoniale perpetrata algoritmicamente, dove l'individuo viene danneggiato nella sua capacità come agente epistemico di fornire conoscenza su sé stesso e sulla propria identità epistemica e pratica. Infatti, nonostante possibili sforzi dell'individuo al fine di farsi percepire nella propria singolarità e irriducibilità alle categorie algoritmiche citate, qualora informato del *modus operandi* descritto, le tecnologie algoritmiche, per come ora progettate, finirebbero comunque per ignorare o fornire minor peso alla prospettiva del singolo in virtù della relativa rappresentazione stereotipata (*baised*) poiché quantitativamente più ricorrente. Infatti, come emerso nei casi di discriminazioni algoritmiche discussi nel dibattito in materia, i sistemi algoritmici finiscono per stereotipare l'individuo, trattandolo come un mero aggregato di associazioni ricorrenti inferibili dai dati a disposizione, e privandolo, in parallelo, della possibilità di contribuire alle pratiche epistemiche che pure lo coinvolgono (quali la ridefinizione dell'ambiente epistemico a cui è esposto e di elementi epistemici personali quali le credenze) in modo *univoco*, cioè sulla base della persona particolare (informata da relazioni, affiliazioni, valori, affetti, impegni e progetti particolari) che è.

Le tecnologie algoritmiche, allo stato attuale, appaiono oggi dunque capaci anche di danneggiare gli individui in quanto soggetti epistemici nella misura in cui privano loro di partecipare attivamente ai processi di produzione di conoscenza, o alle pratiche epistemiche che li riguardano in senso ampio, in modo duplice.

Ingiustizia  
epistemica

Call for papers:  
"Intelligenza  
Artificiale:  
prospettive  
bioetiche,  
biogiuridiche e  
sociali"

Volume 9 ■ 2024

theFuture  
ofScience  
andEthics

19

In primo luogo, gli individui risultano danneggiati *by design* nella loro posizione epistemica in quanto posti in condizioni di mancanza di risorse epistemiche collettive (svantaggio epistemico) per comprendere i processi e gli annessi fenomeni in cui sono coinvolti e le decisioni di matrice algoritmica a cui sono soggetti (*ingiustizia ermeneutica*). In secondo luogo, in quanto sistematicamente declassati o esclusi quali valide fonti di conoscenza nei processi algoritmici in questione a favore di una loro rappresentazione e trattamento da parte di tali sistemi perlopiù basati su correlazioni semplificanti, distorte o errate, cioè stereotipi e pregiudizi identitari ingiusti (*ingiustizia testimoniale*).

#### **4. TECNOLOGIE EPISTEMICHE E SFIDE MORALI: IL RUOLO DELLA FILOSOFIA MORALE NELL'ETICA DELL'IA**

L'analisi etica condotta ha mostrato come le tecnologie algoritmiche *qua* tecnologie intrinsecamente epistemiche possano non solo promuovere la nostra agency epistemica ma anche danneggiarla, creando danni epistemicamente *ingiusti* riconducibili a vere e proprie forme di ingiustizia epistemica testimoniale ed ermeneutica. Questi danni sono moralmente carichi: pongono *sfide morali serie*, già ravvisabili nei costi pratici discussi sopra, che appare qui fondamentale almeno menzionare.

L'opacità (su due livelli: tecnica e politica) nella progettazione di tali sistemi non consente all'individuo di verificare i processi algoritmici a cui è soggetto, richiederne giustificazione, eventualmente contestarli, o chiederne la modifica; in tal senso, riducendo la nostra agency epistemica, minacciano anche la nostra autonomia morale e agency *pratica*, la nostra capacità di avallare criticamente gli *output* algoritmici e il portato epistemico e valoriale che questi riflettono e/o incorporano. A essere sempre più esiguo è, dunque, anche il controllo che possiamo più o meno esercitare su ciò che informa, plasma o persino determina i nostri processi decisionali, le nostre scelte e azioni, e degli elementi che li motivano, quali preferenze, convinzioni, valori, relazioni, che appaiono sempre più plasmati dai sistemi algoritmici; a essere messa in discussione è anche la relativa autenticità, la pluralità epistemica e assiologica e la riflessione critica di cui necessitano per una formazione e adesione personale realmente genuina. A essere sfidata è anche la nostra possibilità

di autorialità nei processi di formazione delle nostre identità personali come persone particolari, e la libertà di orientarle o improntarne lo sviluppo sulla base di ciò che per noi *qua* agenti epistemicamente e moralmente singolari è significativo e realmente (non algoritmicamente) conta.

L'analisi elaborata evidenzia, quindi, il ruolo fondamentale della filosofia e dell'epistemologia morale all'interno della teoria etica applicata alle IA come capaci di fornire gli strumenti concettuali necessari all'indagine non solo delle ingiustizie sociali riproducibili dai sistemi di IA, su cui vi è già un dibattito fervente, ma anche dei danni epistemicamente connessi a forme particolari di ingiustizia epistemica concatenati alle caratteristiche dei sistemi in questione come attualmente progettati. Tali strumenti appaiono altresì cruciali per facilitare lo sviluppo di criteri di *ethics by design*, cioè di quei requisiti etici che debbono orientare la progettazione dei sistemi di IA tenendo conto dei rischi messi in luce, per la realizzazione di tecnologie *epistemicamente* (al fine di essere anche socialmente) giuste, cioè capaci di rispettare gli individui *qua personae* a partire dal loro rispetto quali agenti epistemicamente.

#### **NOTE**

1. Julian Savulescu e Hannah Maslen, "Moral Enhancement and Artificial Intelligence: Moral AI?", in J. Romportl, E. Zackova, J. Kelemen, *Beyond Artificial Intelligence. Topics in Intelligent Engineering and Informatics*, vol. 9, Springer, Cham, 2015.
2. Safiya Umoja Noble, *Algorithms of oppression: How search engines reinforce racism* (New York: New York University Press, 2018).
3. Per un'introduzione al tema soprattutto in relazione a questioni di privacy mentale si veda Fiorella Battaglia, "Algoritmi predittivi e ingiustizia epistemica", in M. Galletti, S. Zipoli Caiani, *Filosofia dell'Intelligenza Artificiale. Sfide etiche e teoriche*, Bologna: Il Mulino, 2024, 63-82.
4. Miranda Fricker, *Epistemic Injustice: Power and the Ethics of Knowing* (Oxford: Oxford University Press, 2007).
5. Simona Tiribelli, *Moral Freedom in the Age of Artificial Intelligence*, Milan-London: Mimesis International, 2022.

6. Paul Humphreys, *Extending ourselves: Computational science, empiricism, and scientific method*, Oxford: Oxford University Press, 2004.

7. Ramon Alvarado, "AI as an Epistemic Technology", *Science and Engineering Ethics* 29, 32 (2023).

8. Fricker 2007, *Epistemic Injustice*, cit., p. 1.

9. Gayle Pohlhaus, "Discerning the Primary Epistemic Harm in Cases of Testimonial Injustice", *Social Epistemology* 28(2), 9, 2014.

10. Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Cambridge (MA): Harvard University Press, 2015.

11. Simona Tiribelli, *Identità personale e algoritmi. Una questione di filosofia morale*, Roma: Carocci editore, 2023.

12. Sábèlo Mhlambi, Simona Tiribelli, "Decolonizing AI Ethics. Relational Autonomy to Counter AI Harms", *To-poi*, 2023.

13. Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future and The New Frontier of Power*, New York: Public Affairs, 2019.

Ingiustizia  
epistemica

Call for papers:  
"Intelligenza  
Artificiale:  
prospettive  
bioetiche,  
biogiuridiche e  
sociali"

Volume 9 ■ 2024

theFuture  
ofScience  
andEthics

21