prospettiveinORGANIZZAZIONE
Rivista trimestrale di Organizzazione Aziendale
ISSN: 2465-1753

COMPRENDERE I BIAS DELL'INTELLIGENZA ARTIFICIALE: RIFLESSIONI DA UNA PROSPETTIVA IMPRENDITORIALE

Pubblicato il12 Dicembre 2025 di Smacchia Marco, Cipriano Michele e Za Stefano

Categoria: Organizzazione: Teorie e Progettazione

Marco Smacchia[1], Michele Cipriano[2], Stefano Za[1]

[1] *Università degli Studi "G. d'Annunzio" Chieti-Pescara*

[2] *Università Cattolica del Sacro Cuore, Piacenza*

Abstract

L'intelligenza artificiale introduce nuovi rischi legati a distorsioni decisionali. Attraverso un'indagine qualitativa tra imprenditori, questo studio evidenzia come i bias, di natura tecnica o sociale, influenzino i processi

organizzativi. Il contributo principale è costituito da uno strumento di supporto manageriale: una matrice che aiuta a valutare le criticità operative e strategiche e a orientare interventi concreti di mitigazione.

Introduzione e motivazione dello studio

Negli ultimi anni, l'intelligenza artificiale (IA) è diventata una componente sempre più centrale dei processi organizzativi, incidendo in modo significativo sulle attività decisionali, operative e strategiche (Huang & Rust, 2018; Rajkomar et al., 2019). Il crescente impiego di soluzioni IA è stato favorito dalla disponibilità di grandi quantità di dati, dal progresso degli algoritmi e da una potenza computazionale sempre più accessibile (Collins et al., 2021). Tuttavia, se da un lato tali tecnologie offrono grandi opportunità, dall'altro sollevano questioni etiche e organizzative di crescente rilevanza, in particolare legate alla presenza di bias, intesi come distorsioni nei risultati di soluzioni IA che favoriscono alcuni individui o gruppi a discapito di altri, imputabili a fattori di natura sociale o computazionale (Akter et al., 2022; Schwartz, 2022).

La letteratura ha ampiamente evidenziato come l'IA, se non adeguatamente progettata, sviluppata e monitorata, possa generare decisioni discriminatorie o distorte, rafforzando disuguaglianze esistenti o creandone di nuove (Mehrabi et al., 2021). Questi bias possono manifestarsi in diverse fasi del ciclo di vita dell'IA, dalla progettazione, all'addestramento di questi modelli, fino all'implementazione, e possono essere causati da fattori tecnici, come l'incompletezza dei dati o l'uso di metodi inadatti, o da elementi sociali e culturali insiti nei dati o nei processi decisionali (Akter et al., 2022).

Negli ultimi anni, l'approccio alla questione dei bias si è evoluto, passando da soluzioni meramente tecniche a visioni più ampie, che includono dimensioni etiche, organizzative e sociotecniche (Akter et al., 2021; Dignum, 2018). Tale prospettiva riconosce che i sistemi IA non possiedono una componente puramente tecnica, ma sono inseriti in contesti organizzativi complessi, influenzati da valori, comportamenti e interazioni tra persone e tecnologie (Sarker et al., 2019). In questo senso, affrontare i bias richiede una comprensione integrata degli aspetti computazionali e sociali che influenzano il funzionamento e l'impatto dei sistemi di IA.

Un utile riferimento teorico per analizzare la natura dei bias è fornito dal framework del National Institute of Standards and Technology (Schwartz, 2022), che distingue tre categorie principali: *bias sistemico*, legato a pratiche sociali e istituzionali consolidate; *bias computazionale*, connesso ai limiti tecnici dei dati e degli algoritmi; e *bias umano*, dovuto a percezioni soggettive e processi cognitivi. Questa classificazione è coerente con le tradizioni sociotecniche, che pongono al centro l'interazione tra persone, tecnologie e contesto.

Tuttavia, la letteratura evidenzia un divario tra riflessioni teoriche e indagini empiriche. Se da un lato esistono molti studi concettuali sul tema, vi è ancora una carenza di evidenze basate su esperienze dirette di chi progetta e implementa sistemi di IA nelle organizzazioni, soprattutto in contesti imprenditoriali e di piccole e medie imprese (PMI). In particolare, mancano approfondimenti su come i bias vengano percepiti da chi opera

quotidianamente con l'IA, e su quali strategie vengano adottate (o del tutto trascurate), per mitigarli (Kordzadeh & Ghasemaghaei, 2021).

In questo quadro si inserisce il presente contributo, che si propone di esplorare il modo in cui gli imprenditori percepiscono i bias nei sistemi di IA progettati e sviluppati all'interno delle proprie aziende. L'obiettivo è duplice: (i) comprendere come tali bias vengano riconosciuti e classificati; (ii) analizzare come le decisioni organizzative incidano sulla manifestazione e sulla gestione dei bias in contesti reali.

Per farlo, adottiamo una prospettiva qualitativa, basata su uno studio di casi multipli condotto attraverso interviste semi-strutturate a imprenditori italiani attivi nel settore dello sviluppo di soluzioni IA. Questa impostazione consente di cogliere le sfumature percettive e decisionali, integrando elementi tecnici e socio-organizzativi, e fornendo una lente analitica utile sia per la riflessione teorica sia per il supporto alla pratica manageriale.

Metodologia della ricerca

Questo studio adotta un approccio qualitativo, volto a esplorare in profondità la percezione e la comprensione del bias dell'IA da parte di imprenditori attivi nello sviluppo di soluzioni digitali. Per analizzare il fenomeno, si è scelto di impiegare un disegno di studio di casi multipli (Yin, 2018), ritenuto particolarmente adeguato ad analizzare fenomeni complessi in contesti reali, quando i confini tra oggetto di studio e contesto sono poco definiti.

Selezione dei Casi e Raccolta Dati

Il processo di raccolta dati si è basato su interviste semistrutturate condotte con imprenditori e figure chiave coinvolte nella progettazione e nell'implementazione di sistemi di IA all'interno delle rispettive imprese. Il campione include due aziende italiane che operano in ambiti diversi (es. servizi, tecnologia, consulenza), selezionate attraverso un campionamento intenzionale basato su due criteri: (i) presenza di soluzioni IA progettate internamente; (ii) coinvolgimento diretto del gruppo imprenditoriale nella definizione delle funzionalità IA.

Le interviste sono state condotte tra marzo e giugno 2022 e hanno avuto una durata media di circa 60 minuti. Tutte le conversazioni sono state registrate, trascritte integralmente e sottoposte a verifica da parte degli intervistati per garantirne l'accuratezza. Il protocollo di intervista includeva domande aperte su: (i.) la natura delle soluzioni IA adottate; (ii.) i processi decisionali relativi alla loro progettazione; (iii.) la consapevolezza e gestione dei potenziali bias; (iv.) la percezione delle implicazioni etiche e organizzative.

Analisi dei Dati

Le trascrizioni sono state analizzate tramite codifica tematica (Braun & Clarke, 2006), un metodo flessibile ma rigoroso per individuare, analizzare e interpretare schemi ricorrenti all'interno di dati qualitativi. L'analisi si è sviluppata in più fasi:

1. lettura esplorativa delle trascrizioni;
2. generazione dei codici iniziali in modo induttivo;
3. raggruppamento dei codici in categorie tematiche;
4. confronto tra ricercatori per affinare e validare i temi emersi.

Dall'analisi sono emerse 13 codifiche distinte, successivamente raggruppate in due categorie principali: bias computazionale, riferito a limiti tecnici come dati incompleti o modelli non rappresentativi, e bias sociale, legato a norme, valori o stereotipi incorporati nei processi organizzativi (Tabella 1). Queste due categorie rappresentano le principali modalità con cui i partecipanti interpretano le distorsioni generate o rinforzate dai sistemi IA. I codici emersi costituiscono inoltre la base per la costruzione interpretativa sviluppata nella sezione successiva.

L'utilizzo combinato di un approccio esplorativo, un campione mirato e un processo di codifica strutturato ha permesso di garantire un buon livello di affidabilità e coerenza metodologica, pur riconoscendo i limiti tipici degli studi qualitativi in termini di generalizzabilità.

Analisi e risultati

L'analisi delle interviste ha evidenziato due modalità principali attraverso cui gli imprenditori percepiscono il fenomeno del bias nei sistemi di IA: una lettura tecnica/computazionale e una lettura sociale/sistemica. Queste due dimensioni rappresentano i filoni interpretativi predominanti nel modo in cui gli intervistati comprendono e affrontano le distorsioni generate o amplificate dalle tecnologie IA all'interno dei loro contesti organizzativi.

Bias Computazionale: Una Questione di Dati e Modelli

La prima interpretazione, maggiormente diffusa, attribuisce la presenza di bias a cause di natura tecnica, legate principalmente ai dati utilizzati per addestrare gli algoritmi, alla qualità dei modelli e alla progettazione delle funzionalità IA. Questo approccio si fonda sull'idea che il bias derivi da errori, mancanze o limitazioni nella base informativa o nella logica computazionale del sistema.

Molti imprenditori hanno evidenziato come il dataset rappresenti la componente più critica nella prevenzione dei bias. In diversi casi, la difficoltà di reperire dati bilanciati e rappresentativi è stata descritta come una sfida quotidiana:

"Abbiamo cercato di evitare distorsioni nei risultati usando dataset pubblici, ma spesso ci accorgiamo solo in fase di test che mancano categorie o profili fondamentali". (Top Manager, azienda A)

"Per certi clienti, i dati disponibili sono semplicemente troppo pochi o troppo simili: il sistema finisce per imparare da un'unica prospettiva". (Top Manager, azienda A)

Oltre alla scarsità o sbilanciamento dei dati, alcuni intervistati hanno sottolineato anche i limiti degli algoritmi stessi nel trattare scenari complessi o ambigui. È emersa una certa consapevolezza circa il rischio che modelli "black-box" introducano comportamenti opachi e difficilmente verificabili:

"La rete neurale funziona, ma non sappiamo esattamente cosa la spinge a scegliere un risultato piuttosto che un altro. E questo è un problema se qualcosa va storto." (Top Manager, azienda B)

In generale, i bias computazionali sono percepiti come problematiche tecniche risolvibili attraverso una migliore gestione del ciclo di vita dell'IA: raccolta dati più robusta, validazione trasparente dei modelli e aggiornamenti continui. Tuttavia, in pochi casi sono state segnalate procedure sistematiche per gestire questi aspetti, e la maggior parte delle azioni è risultata reattiva e informale.

Bias Sociale: Valori Impliciti e Pratiche Organizzative

Una seconda interpretazione, meno comune ma concettualmente più matura, riconosce che i sistemi di IA riflettono valori, norme e assunzioni sociali incorporati nei dati o nei processi organizzativi. In questa visione, il bias non è solo un problema tecnico, ma un fenomeno socio-organizzativo che coinvolge il modo in cui le tecnologie vengono concepite, addestrate e impiegate nelle pratiche aziendali.

"L'algoritmo è imparziale solo in apparenza, perché dietro c'è sempre una decisione umana su cosa considerare 'normale' o 'desiderabile'." (Top Manager, azienda B)

“A volte ci accorgiamo che certe funzioni riflettono il modo di pensare del gruppo di sviluppo, anche inconsciamente. Non è neutrale.” (Top Manager, azienda B)

Un altro intervistato ha sottolineato come anche le dinamiche interne al gruppo possano influenzare le scelte progettuali:

“Chi scrive l’algoritmo ha un impatto enorme. Magari senza accorgersene porta dentro il proprio modo di vedere il mondo.” (Top manager, azienda A)

Questa prospettiva mette in discussione l’idea che basti “ripulire i dati” per eliminare il bias, sottolineando piuttosto l’importanza del contesto sociale e culturale in cui la tecnologia viene sviluppata. Alcuni partecipanti hanno evidenziato che i criteri di ottimizzazione algoritmica spesso escludono aspetti etici o relazionali importanti:

“Quando sviluppi un algoritmo per suggerire profili lavorativi, stai facendo delle scelte su cosa è un ‘buon’ candidato. Non è solo una questione tecnica.” (Top manager, azienda B)

Categoria	Codici principali	Esempi associati
Bias computazionale	Dati incompleti, dataset sbilanciati, opacità algoritmica	Dataset pubblici insufficienti; modelli “black box”
Bias sociale	Assunzioni implicite, valori del team, stereotipi normativi	Criteri di selezione impliciti; visione del team

Tabella 1. Riepilogo Sintetico dei Codici Emersi dall’Analisi Tematica.

Tuttavia, questa consapevolezza non sempre si traduce in pratiche operative consolidate. Sebbene alcuni imprenditori abbiano introdotto meccanismi di validazione etica informale o confronto interno, nessuna delle aziende intervistate ha dichiarato di adottare framework strutturati di auditing o fairness assessment.

Nel complesso, i risultati evidenziano una forte dicotomia interpretativa. Da un lato, una visione tecnicista che tende a trattare il bias come una conseguenza non voluta di scelte algoritmiche migliorabili. Dall’altro, una lettura critica e sistemica, che riconosce le implicazioni più profonde delle decisioni incorporate nei sistemi IA. Le due visioni non si escludono, ma coesistono spesso in tensione, e rispecchiano il livello di maturità e riflessività dei soggetti coinvolti.

Queste evidenze costituiscono la base empirica per la riflessione sviluppata nella sezione successiva, dove verrà proposta una matrice interpretativa in grado di rappresentare le diverse configurazioni di bias e il loro impatto

decisionale nelle organizzazioni.

Discussioni

Sulla base dei risultati emersi dall'analisi qualitativa, è stato sviluppato un framework concettuale che consente di interpretare e categorizzare i bias dell'IA in relazione al loro tipo e al livello decisionale in cui si manifestano. L'obiettivo è fornire uno strumento utile per comprendere non solo che tipo di bias si verifichi, ma anche dove esso incida nel contesto organizzativo.

La Matrice: Tipo Di Bias E Livello Decisionale

La **matrice a quattro quadranti** proposta (Figura 1) incrocia due dimensioni principali:

- l'origine del bias, distinta tra *computazionale* (tecnico) e *sociale* (sistemico);
- il livello decisionale su cui il bias ha impatto, che può essere *strategico* (es. definizione di mission, modelli di business, scelte di mercato) o *operativo* (es. selezione di clienti, classificazione dati, interazioni uomo-macchina).

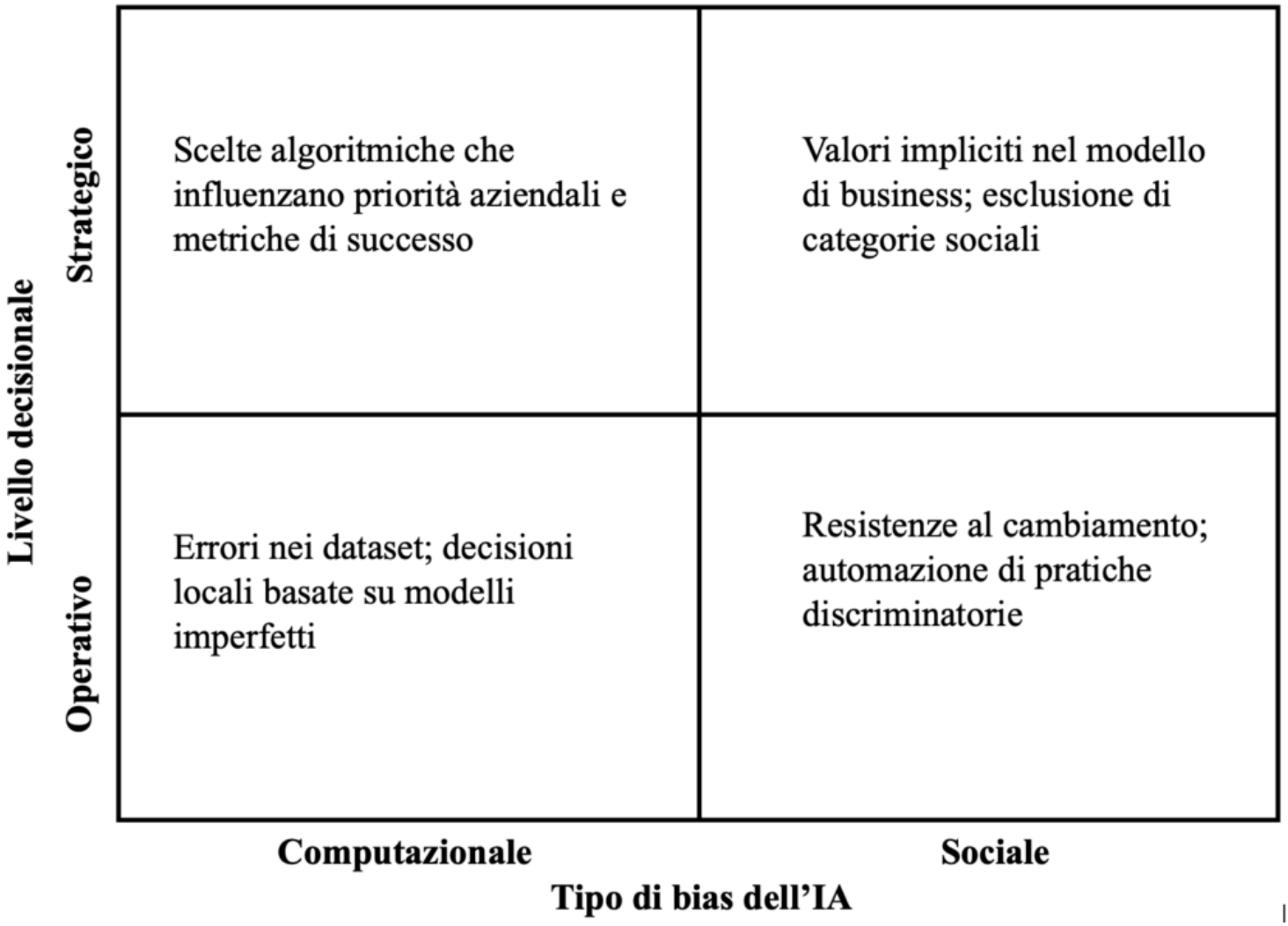

Figura 1. Matrice Interpretativa Dei Bias IA In Ambito Organizzativo

Interpretazione dei Quadranti della Matrice

Quadrante 1. Bias computazionale - livello operativo: questo è il quadrante più frequentemente menzionato dagli intervistati. Riguarda errori o distorsioni dovuti a dataset incompleti, algoritmi opachi o modelli mal calibrati che influenzano le attività quotidiane (es. errori nelle predizioni sulla manutenzione dei macchinari). È spesso affrontato con soluzioni tecniche, ma raramente con protocolli formali.

"Ci siamo accorti che l'algoritmo escludeva certi errori solo perché il training set non ne aveva abbastanza." (Top Manager, azienda A)

Quadrante 2. Bias computazionale - livello strategico: meno visibile ma altamente impattante, si manifesta quando i parametri computazionali guidano scelte strategiche (es. priorità nei prodotti, target di mercato). Gli imprenditori riconoscono che le metriche definite a livello algoritmico possono indirizzare l'impresa verso direzioni non previste o eticamente discutibili.

"Abbiamo scelto un criterio predittivo che favoriva i clienti già forti, ma poi ci siamo resi conto che stava snaturando la nostra offerta." (Top Manager, azienda A)

Quadrante 3. Bias sociale - livello operativo: riguarda pratiche quotidiane che incorporano stereotipi o pregiudizi non consapevoli, spesso trasferiti nei processi automatizzati e che possono comportare la nascita di resistenze da parte degli utilizzatori. Ad esempio, un algoritmo di screening dei CV può perpetuare preferenze implicite presenti nei dati storici dell'azienda.

"Il sistema suggeriva sempre gli stessi profili, come se avesse imparato dai vecchi pregiudizi di chi faceva selezione." (Top manager, azienda B)

Quadrante 4. Bias sociale - livello strategico: è il quadrante più trascurato, ma potenzialmente il più critico. Riguarda i valori e le assunzioni profonde che guidano la progettazione stessa delle soluzioni IA. Alcuni imprenditori hanno riconosciuto che la cultura del team o il contesto sociale possono influenzare la visione del problema e del suo "miglior" trattamento.

"Scegliere cosa ottimizzare non è mai neutro. Abbiamo fatto delle scelte che oggi rivedremmo, perché ci siamo resi conto che escludevano alcune categorie." (Top Manager, azienda B)

Implicazioni Manageriali

Il framework sviluppato fornisce una mappa concettuale utile ai decisori per identificare dove agiscono i bias e quali leve organizzative possono essere attivate. In particolare, emergono tre implicazioni pratiche principali:

1. **Consapevolezza distribuita**: i bias non riguardano solo gli sviluppatori, ma coinvolgono tutta l'organizzazione. È necessario promuovere una cultura diffusa di responsabilità algoritmica.

2. **Audit multilivello**: servono strumenti che valutino gli impatti sia tecnici che sociali dell'IA, a diversi livelli decisionali. Non basta validare il modello, occorre interrogarsi sulle sue finalità.

3. **Inclusione di prospettive**: la diversità nei gruppi di lavoro e il coinvolgimento di stakeholder eterogenei aumentano la probabilità di individuare i bias sociali nascosti e di evitarne la cristallizzazione.

Il framework può essere utilizzato non solo in fase di progettazione, ma anche per la revisione critica di sistemi esistenti, aiutando le imprese a riconoscere dinamiche di esclusione non intenzionali e ad attuare strategie correttive orientate all'equità e alla trasparenza.

Conclusioni e Implicazioni per la teoria e la pratica

Il presente contributo ha esplorato la percezione del bias algoritmico da parte di imprenditori coinvolti nello sviluppo di soluzioni di IA, mettendo in luce come il fenomeno venga interpretato attraverso due principali punti di vista: uno tecnico, focalizzato sui dati e sugli algoritmi, e uno sociale, legato a valori, norme e pratiche organizzative. L'analisi qualitativa ha permesso di cogliere la varietà di esperienze e la complessità delle sfide affrontate, evidenziando come il bias non sia un problema esclusivamente tecnico, ma un fenomeno sociotecnico, profondamente radicato nei contesti decisionali.

La matrice proposta nella sezione precedente consente di sistematizzare tali evidenze e offre uno strumento concettuale utile per l'analisi e la gestione del bias nei diversi livelli dell'organizzazione. In particolare, la distinzione tra bias computazionale e bias sociale, unita alla separazione tra livello strategico e operativo, permette ai decisori di identificare con maggiore precisione le aree di vulnerabilità e di definire interventi mirati.

Implicazioni Manageriali

I risultati di questo studio suggeriscono diverse implicazioni concrete per imprenditori, manager e progettisti di sistemi IA, in particolare per coloro che operano in contesti organizzativi agili e orientati all'innovazione.

Anzitutto, è fondamentale che le organizzazioni imparino ad andare oltre una visione puramente tecnica del bias, che tende a trattare le distorsioni come semplici problemi di pulizia dei dati o di ottimizzazione algoritmica. Al contrario, è necessario riconoscere che molte distorsioni hanno radici più profonde, legate ai valori organizzativi, alle priorità strategiche e alle pratiche operative. Questo richiede uno sforzo consapevole per integrare l'analisi etica e sociale fin dalle fasi iniziali della progettazione dei sistemi.

In secondo luogo, le imprese dovrebbero dotarsi di procedure strutturate di audit etico e riflessione critica, anche informali, che permettano di valutare periodicamente l'impatto delle soluzioni di IA implementate. Tali momenti possono assumere la forma di "ethical checkpoints" nei progetti, workshop interni, o consultazioni con soggetti esterni qualificati. L'obiettivo è garantire che le decisioni algoritmiche siano coerenti con i valori dichiarati dell'organizzazione e che non producano effetti indesiderati su stakeholder interni o esterni.

Un ulteriore elemento chiave è l'adozione di un approccio multidisciplinare e partecipativo nella progettazione dei sistemi. Coinvolgere diverse funzioni aziendali (dai data scientist agli esperti di dominio, fino ai responsabili delle risorse umane o agli specialisti di compliance) consente di aumentare la sensibilità verso possibili bias impliciti e di favorire soluzioni più robuste, sostenibili e inclusive.

Infine, emerge con forza l'esigenza di promuovere una cultura diffusa della responsabilità algoritmica, in cui i sistemi IA siano percepiti non come soggetti autonomi, ma come strumenti le cui decisioni devono essere continuamente interrogate, interpretate e, se necessario, corrette. Questa cultura si fonda sulla trasparenza, sul dialogo e sull'accountability e può rappresentare un vantaggio competitivo per quelle organizzazioni che sapranno integrare l'etica nella propria capacità innovativa.

Limiti e sviluppi futuri

Come ogni studio di natura qualitativa, anche il presente lavoro presenta alcune limitazioni che vanno considerate nel valutare la portata dei risultati. In primo luogo, il numero ridotto di casi analizzati e il focus geografico circoscritto al contesto italiano limitano la possibilità di generalizzare i risultati ad altri settori o paesi. Tuttavia, l'approccio interpretativo adottato punta a far emergere pattern concettuali e riflessioni utili anche oltre il campione studiato.

In secondo luogo, le evidenze raccolte si basano su interviste narrative auto-riferite, il che implica il rischio di bias di desiderabilità sociale o inconsapevolezza da parte degli intervistati rispetto ad alcuni aspetti problematici. Nonostante gli accorgimenti metodologici adottati (es. triangolazione interna, validazione delle trascrizioni), questo elemento rappresenta un limite tipico ma rilevante degli studi esplorativi.

Per quanto riguarda gli sviluppi futuri, sarebbe utile estendere l'indagine ad altri contesti organizzativi, come grandi imprese, enti pubblici o organizzazioni internazionali, per confrontare diverse culture aziendali nell'approccio al bias algoritmico. Inoltre, l'integrazione tra metodi qualitativi e quantitativi (ad esempio attraverso survey strutturate su larga scala) potrebbe contribuire a verificare la ricorrenza delle dinamiche emerse.

Infine, un importante sviluppo applicativo potrebbe essere rappresentato dalla progettazione di toolkit operativi, come checklist, canvas o linee guida, pensati per supportare imprenditori e manager nell'identificazione precoce dei bias e nella costruzione di soluzioni IA più eque e consapevoli.

Bibliografia

Akter, S., McCarthy, G., Sajib, S., Michael, K., Dwivedi, Y. K., D'Ambra, J., & Shen, K. N. (2021). Algorithmic bias in data-driven innovation in the age of AI. In International Journal of Information Management (Vol. 60). Elsevier Ltd. https://doi.org/10.1016/j.ijinfomgt.2021.102387

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative research in psychology, 3(2), 77-101.

Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. International Journal of Information Management, 60(June), 102383. https://doi.org/10.1016/j.ijinfomgt.2021.102383

Dignum, V. (2018). Ethics in artificial intelligence: introduction to the special issue. Ethics and Information Technology, 20(1), 1–3. https://doi.org/10.1007/s10676-018-9450-z

Huang, M. H., & Rust, R. T. (2018). Artificial Intelligence in Service. Journal of Service Research, 21(2), 155–172. https://doi.org/10.1177/1094670517752459

Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: review, synthesis, and future research directions. European Journal of Information Systems, 31(3), 388–409. https://doi.org/10.1080/0960085X.2021.1927212

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. ACM Computing Surveys, 54(6). https://doi.org/10.1145/3457607

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347-1358.

Sarker, S., Chatterjee, S., Xiao, X., & Elbanna, A. (2019). The Sociotechnical Axis of Cohesion for the IS Discipline: Its Historical Legacy and its Continued Relevance. MIS Quarterly, 43(3), 695–719. https://doi.org/10.25300/MISQ/2019/13747

Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., & Hall, P. (2022). Towards a standard for identifying and managing bias in artificial intelligence. US Department of Commerce, National Institute of Standards and Technology, 3. https://doi.org/10.6028/NIST.SP.1270

Yin, R. K. (2018). Case Study Research and Applications (Sixth Edition). SAGE.