2025

SICUREZZA E SCIENZE SOCIALI

ANNO XIII
N. 2BIS/2025

Le intelligenze artificiali: verso dove?

a cura di
Roberto Cipriani, Sara Sbaragli

ISSN 2283-8740, ISSNE 2283-7523
OPEN ACCESS

SICUREZZA E SCIENZE SOCIALI

La rivista esce sotto l'alto patrocinio
dell'Università degli Studi di Perugia

Con il patrocinio del
Comune di Narni

La rivista si propone di sostenere e di dare voce alle esigenze e alle istanze pluralistiche dei Corsi di laurea universitari che, nel contesto italiano, affrontano in maniera specifica le tematiche di carattere criminologico.

Alma Mater Studiorum – Università di Bologna
Laurea Magistrale in "Scienze criminologiche per l'investigazione e la sicurezza"

Università degli Studi Magna Græcia di Catanzaro
Laurea Magistrale "Organizzazioni e mutamento sociale"

Università Cattolica del Sacro Cuore
Laurea Magistrale in "Politiche pubbliche - curriculum Politiche per la sicurezza"

Università degli Studi "G. d'Annunzio" Chieti – Pescara
- Laurea Triennale "Sociologia e criminologia"
- Laurea Magistrale "Ricerca Sociale, politiche della sicurezza e criminologia"

Università degli Studi di Perugia
- Laurea Triennale "Scienze per l'investigazione e la sicurezza"
- Laurea Magistrale "Scienze Socio-antropologiche per l'integrazione e la sicurezza sociale"

SICUREZZA E SCIENZE SOCIALI

Sommario

Introduzione.
Le intelligenze artificiali: verso dove?

di *Costantino Cipolla*[*], *Sara Sbaragli*[**]

Scrivere in poche righe qualcosa di sensato sull'intelligenza artificiale, il mostro della nostra era, in chiave sociologica è un'impresa quasi impossibile. Intanto, le intelligenze tecnologiche sono tantissime e coprono ogni ambito, ormai, del nostro vivere personale e sociale. Esse sono artificiali in quanto non prodotte direttamente dalla natura umana, ma dalla sua intelligenza cerebrale, empirica, storica e così via. Ed anche per questa via esse sono un'infinità e circondano e influenzano per ogni dove (spesso inconsciamente) le nostre scelte e il nostro stesso modo di concepire e di muoversi nel mondo sociale. Nelle società digitali, evolute o meno, la sicurezza relazionale è diventata uno dei fattori ritenuti più qualificanti della convivenza civile. Superata la soglia della necessità, nelle sue molteplici dimensioni, dati per scontati e acquisiti altri aspetti della vita quotidiana, il sentirsi e vivere al sicuro sono diventati un'esigenza basilare e imprescindibile di ogni individuo mediamente ricco e socializzato. Ed in questo mercato in fiorente ascesa e dai mille risvolti poteva non insinuarsi il digitale applicato in molti modi, in difformi circostanze e nei più svariati contesti? La letteratura in generale su questi temi e su quello più mirato che qui ci compete è sterminata, soprattutto in lingua inglese (Schwarz, 2025; Neri, Cordeiro, 2025; Roumbanis, 2025; Romele, 2023). Essa, però, tende a crescere anche nell'ambito della nostra cultura sociologica nazionale, soprattutto da parte delle giovani generazioni, come questo numero della Rivista, con altri, dimostra ampiamente (Cristianini, 2023; Balbi, 2022; Natale, 2022; Grassi, 2020).

La sicurezza sociale, per il vero, riguarda fenomeni anche molto lontani tra di loro, che oggi presentano tutti, innescati e attraversati dall'intelligenza artificiale nelle sue infinite forme (come già scritto), il carattere di una precipua originalità. Il passato in questo caso non serve molto a comprendere il futuro, se non per differenza. Questo non può che essere piuttosto

DOI: 10.5281/zenodo.17522762

[*] Università di Bologna. costantino.cipolla@unibo.it.
[**] Università di Napoli Federico II. sarasbaragli@gmail.com.

imprevedibile, spesso improvviso, quasi sempre in movimento più o meno forsennato. La sicurezza del futuro la si può vedere potente, anticipatrice, a distanza, intellettuale, onnipervasiva e sempre presente, al servizio, quasi in modo costitutivo, di quella umana o governata fisicamente e direttamente dall'uomo, che in questo caso appare sempre più incarnato in una donna, anche, tra tante difficoltà, se non altro per questioni di alleggerimento corporeo. Questa sorta di rivoluzione continua si diffonde in modo diversificato in tanti campi, come si può vedere più oltre, e come ora sintetizzeremo seguendo, con qualche salto e aggregazione, i contenuti del presente numero della rivista.

Gli articoli del numero si concentrano sulle molteplici applicazioni e implicazioni dell'intelligenza artificiale nella società contemporanea, esplorandone tanto le opportunità quanto i rischi, in una varietà di contesti che spaziano dal sanitario al sociale, dall'educativo al religioso, fino all'ambito della sicurezza.

Un primo filone affronta il ruolo dell'IA nell'invecchiamento della popolazione, mettendo in luce come questa tecnologia possa contribuire a contrastare l'isolamento sociale degli anziani e promuovere un invecchiamento attivo e indipendente. In particolare, l'uso dell'intelligenza artificiale nella telemedicina e nelle tecnologie assistive può facilitare l'accesso alle cure, potenziare l'autonomia e rafforzare i legami sociali degli anziani, favorendo una maggiore qualità della vita.

Nel campo della sanità emerge una riflessione critica sulla presenza di pregiudizi razziali nei sistemi di intelligenza artificiale medica. Una revisione sistematica ha evidenziato come l'uso di dati di addestramento non rappresentativi possa perpetuare disparità sanitarie, influenzando negativamente le diagnosi e i trattamenti. Affrontare questi bias richiede interventi correttivi interdisciplinari, capaci di coniugare informatica, etica, medicina e scienze sociali.

Il tema della creatività solleva interrogativi profondi sulla distinzione tra mente umana e artificiale. Viene proposta una riflessione teorica che contrappone la creatività trasformativa dell'essere umano a quella esplorativa della macchina, mettendo in discussione la capacità delle IA di comprendere il contesto, il significato e l'imprevedibilità insite nei processi creativi.

Nel campo del sociale, l'IA viene indagata per il suo potenziale nella prevenzione e nella gestione del fenomeno dei senza dimora. Strumenti predittivi e analisi dei big data possono contribuire a identificare precocemente situazioni di vulnerabilità, ma il loro utilizzo deve essere

bilanciato da un forte presidio etico per evitare forme di discriminazione e marginalizzazione.

In ambito relazionale ed emotivo, si discute l'effetto dell'"*Affective Computing*" e delle chatbot capaci di simulare emozioni umane, con particolare attenzione ai rischi di manipolazione affettiva e alla tendenza degli utenti a sviluppare legami empatici con entità artificiali. L'illusione di una relazione autentica solleva domande sulla progettazione responsabile delle interfacce uomo-macchina.

Un altro contributo analizza la trasformazione della guerra nell'era digitale, introducendo il concetto di "*mosaic warfare*" alimentato dall'IA. Le tecnologie digitali amplificano la vulnerabilità delle infrastrutture critiche e spostano la concezione del conflitto verso forme sempre più astratte e deresponsabilizzanti, dove la distanza tecnologica riduce la percezione del valore della vita umana.

La sfera emotiva viene inoltre approfondita secondo una prospettiva sociologica, mettendo in luce come l'IA possa alterare la gestione delle emozioni e favorire forme di "devianza emozionale", alimentando relazioni fredde e artificiali. L'autenticità delle interazioni umane rischia di essere sostituita da connessioni regolate da algoritmi, con conseguenze sulla coesione sociale.

Si propone poi una lettura simbolico-religiosa dell'IA, che viene interpretata come nuova "religione secolare". Alcune applicazioni di intelligenza artificiale si inseriscono nel contesto spirituale, simulando preghiere e dialoghi con figure religiose. In questo scenario, la tecnica tende a sostituirsi alla trascendenza, offrendo una spiritualità artificiale e istantanea che risponde al bisogno contemporaneo di salvezza senza fede.

Inoltre, i contributi riflettono ulteriormente sulla crescente penetrazione dell'intelligenza artificiale in contesti sempre più diversificati, evidenziandone tanto il potenziale trasformativo quanto le criticità etiche, sociali e operative che ne derivano.

Uno dei temi affrontati riguarda la sicurezza stradale e i rischi legati all'affidabilità dei sistemi digitali installati nei veicoli. Un caso emblematico ha mostrato come un errore tecnico possa portare a falsi allarmi, con implicazioni legali e sociali significative. Si evidenzia la necessità di sviluppare una normativa più precisa e strumenti assicurativi adeguati, in grado di affrontare le conseguenze dei malfunzionamenti dei sistemi automatizzati. A ciò si aggiunge una riflessione sul rischio etico legato all'affidamento crescente a dispositivi automatizzati per decisioni che, se sbagliate, possono mettere in pericolo la vita umana.

Il dibattito si estende al tema della guida autonoma, dove l'uso dell'IA si confronta con i limiti tecnici della sua implementazione su larga scala. Nonostante le aspettative elevate e gli investimenti consistenti, la guida automatizzata rimane confinata a spazi controllati e altamente geolocalizzati. La complessità del traffico reale, la sicurezza dei soggetti più vulnerabili sulla strada e l'incapacità dell'IA di comprendere pienamente il contesto situazionale pongono interrogativi sulla sua reale efficacia nella prevenzione degli incidenti.

Sul fronte militare, l'introduzione dell'intelligenza artificiale nei sistemi di supporto decisionale solleva interrogativi etici profondi. Sebbene queste tecnologie aumentino la rapidità e l'efficacia operativa, vi è il rischio che il personale delegittimi il proprio giudizio in favore delle decisioni automatiche, con possibili conseguenze sul rispetto delle norme internazionali in materia di conflitto armato. Si sottolinea l'importanza di mantenere l'autonomia decisionale umana come elemento centrale, affinché le decisioni restino ancorate ai principi morali e giuridici.

Un altro ambito inedito e controverso riguarda l'integrazione dei sex robot nel lavoro sessuale. L'adozione di queste tecnologie apre interrogativi su privacy, discriminazione e trasformazione della relazione uomo-macchina. Sebbene vengano presentati come strumenti di intrattenimento o alternativa sicura, essi rischiano di rafforzare stereotipi stigmatizzanti e produrre nuove forme di esclusione sociale. Il contributo invita a un approccio critico, fondato su una regolazione attenta e su solidi principi etici.

Vi è un'ampia riflessione filosofica e pedagogica sull'impatto dell'IA nel quotidiano. La paura diffusa verso la tecnologia è attribuita non solo alla sua complessità, ma anche alla percezione di perdita del controllo e dell'esperienza artigiana, dove l'uomo era protagonista diretto del fare. Tuttavia, viene ribadito che la chiave per affrontare questa rivoluzione risiede nell'educazione e nella consapevolezza diffusa. L'umanità deve restare al centro dello sviluppo tecnologico, non come spettatrice, ma come agente attivo capace di guidare l'innovazione lungo un percorso condiviso e responsabile.

Infine, è analizzato il legame tra intelligenza artificiale, capitalismo delle piattaforme e democrazia, mostrando come gli algoritmi non siano neutri ma espressione di interessi economici e politici. Viene messo in luce il ruolo delle *Big Tech* nella concentrazione di potere e nella produzione di disuguaglianze attraverso processi di datificazione e sorveglianza. Il capitalismo digitale, tramite logiche di *nudging* e profilazione, influenza scelte individuali e collettive, minando autonomia e sovranità. In questo quadro, la sfera pubblica rischia di frammentarsi in bolle informative e filtri

algoritmici. L'articolo conclude avvertendo dei rischi per la deliberazione democratica e per la sostanza partecipativa delle società contemporanee.

In sintesi, gli articoli restituiscono un panorama ampio e complesso, in cui l'intelligenza artificiale appare come strumento potente ma ambivalente, capace tanto di migliorare la vita quanto di generare nuove forme di dipendenza, disuguaglianza e disconnessione umana. L'intelligenza artificiale viene osservata come forza potente, da governare con attenzione affinché possa generare benefici collettivi senza erodere diritti, responsabilità e umanità.

Riferimenti bibliografici

Balbi G. (2022). *L'ultima ideologia*. Roma-Bari: Laterza.

Cristianini N. (2023). *La scorciatoia*. Bologna: il Mulino.

Grassi E. (2020). *Etica e intelligenza artificiale*. Roma: Aracne.

Natale S. (2022). *Macchine ingannevoli*. Torino: Einaudi.

Neri H., Cordeiro V. (2025). «Reimagining sociality in the digital age: transcending the interaction/society dichotomy». *Systems Research and Behavioral Science*. https://doi.org/10.1002/sres.3304.

Romele A. (2023). *Digital habitus: a critique of the imaginaries of artificial intelligence*. London-New York: Routledge.

Roumbanis L. (2025). «On algorithmic mediations». *European Journal of Social Theory*. https://doi.org/10.1177/13684310251319677.

Schwarz O. (2025). «The post-choice society: algorithmic prediction and the decentring of choice». *Theory, Culture & Society*. https://doi.org/10.1177/02632764251322062.

Anziani e intelligenza artificiale: quali prospettive e opportunità?

di *Anna Santovito*[*]

L'invecchiamento globale pone sfide sociali, ma l'intelligenza artificiale (IA) può migliorare la qualità della vita degli anziani. Studi recenti (Alessa, Al-Khalifa 2023, Abdollahi *et al.,* 2022) evidenziano come l'IA contrasti la solitudine ed elimini barriere fisiche e sociali. I fattori come pensionamento e ridotta mobilità influiscono sulle relazioni, ma la tecnologia offre soluzioni innovative. L'articolo con una revisione di articoli scientifici sul tema analizza le opportunità e criticità etiche e tecnologiche legate all'IA. Gli obiettivi principali sono comprendere come le tecnologie facilitino l'integrazione sociale e il benessere degli anziani basandosi su studi e ricerche sul tema degli anziani in relazione ai temi della socializzazione, mobilità e telemedicina.

Parole chiave: anziani; intelligenza artificiale; tecnologie; mobilità; socializzazione; sanità.

Elderly people and artificial intelligence: what are the prospects and opportunities?

Global aging poses social challenges, but artificial intelligence (AI) can improve the quality of life of older adults. Recent studies (Alessa, Al-Khalifa 2023, Abdollahi *et al.,* 2022) highlight how AI combats loneliness and eliminates physical and social barriers. Factors such as retirement and reduced mobility affect relationships, but technology offers innovative solutions. The article, with a review of scientific articles on the topic, analyzes the ethical and technological opportunities and criticisms related to AI. The main objectives are to understand how technologies facilitate the social integration and well-being of the elderly by exploring applications in fields such as socialization, mobility and telemedicine.

Keywords: elderly people; artificial intelligence; technologies; mobility; socialisation; healthcare.

Introduzione

Nei paesi industrializzati, l'aumento della popolazione anziana richiede una valutazione attenta delle conseguenze sul sistema di welfare e delle

DOI: DOI 10.5281/zenodo.17523817

[*] Università degli Studi di Bari. santovito.an@gmail.com.

prospettive per affrontare le sfide che sono legate all'aumento dell'aspettativa di vita ed alla diminuzione della natalità (ISTAT, 2020).

A questo proposito, l'invecchiamento della popolazione, lo sviluppo delle tecnologie e l'utilizzo dell'intelligenza artificiale stanno modificando profondamente la società attuale.

In particolare, l'intelligenza artificiale (d'ora in poi IA) si pone come strumento di supporto per la fragilità e la disabilità degli anziani, offrendo soluzioni innovative nel campo della salute, della mobilità e per favorire l'inclusione sociale (Chen, 2020; Cornwell, Waite, 2009; Swami *et al.,* 2007). Secondo Baltes *et al.* (1980), durante il processo di invecchiamento, gli anziani devono affrontare il declino delle risorse biologiche e il supporto offerto dall'ambiente sociale.

L'IA offre soluzioni per alleggerire questo problema, promuovendo l'invecchiamento attivo (WHO, 2002) attraverso strumenti digitali che facilitano la comunicazione e l'accesso a servizi essenziali.

1. L'intelligenza artificiale come ponte per la socializzazione degli anziani

Durkheim (1895) evidenziava che la partecipazione e l'integrazione sociale sono fondamentali per il benessere individuale e collettivo.

La partecipazione sociale implica l'adesione alle norme, l'interazione nelle istituzioni e il coinvolgimento comunitario, sviluppandosi lungo tutto l'arco della vita. Secondo Havighurst (1963), per un invecchiamento di successo è fondamentale restare attivi a livello fisico, mentale e sociale.

La sua teoria dell'attività evidenzia che mantenere un buon livello di autonomia, accettare i cambiamenti fisici e continuare a impegnarsi in relazioni e attività sociali aiuta le persone a vivere meglio la vecchiaia.

In relazione a ciò, la teoria socio-emotiva della selettività di Laura Carstensen (1992) suggerisce che, con l'avanzare dell'età, gli individui tendono a selezionare e concentrarsi su relazioni sociali emotivamente gratificanti e significative.

Questa teoria riporta e riflette una "selettività emotiva" nella scelta delle relazioni per massimizzare il benessere emotivo nel contesto delle tecnologie e per usare le piattaforme digitali nel connettersi con persone con cui hanno legami affettivi.

Lo studio di Chen (2020) analizza l'impatto dell'uso degli smartphone sulla qualità delle relazioni e sul benessere soggettivo degli anziani attraverso una prospettiva del corso della vita. La metodologia adottata prevede l'utilizzo di questionari somministrati a un campione di partecipanti

di diverse età mentre i questionari includevano l'uso degli smartphone, la qualità delle relazioni interpersonali e il benessere soggettivo.

I risultati indicano che l'uso equilibrato e consapevole delle comunicazioni multimediali tramite gli smartphone, come i messaggi di testo e le videochiamate, è associato ad una maggiore qualità delle relazioni e ad un aumento del benessere soggettivo mentre, un uso eccessivo degli smartphone o una comunicazione superficiale possono contribuire a conflitti relazionali e a un senso di isolamento, diminuendo il benessere psicologico.

Uno dei principali ostacoli che gli anziani affrontano nell'uso dell'intelligenza artificiale è il "divario digitale", dovuto alla loro formazione in un contesto analogico, a differenza delle giovani generazioni cresciute in ambienti tecnologici.

Nel dibattito sull'inclusione digitale e sociale, il rapporto tra le generazioni gioca un ruolo chiave, in questo contesto, ricordiamo la teoria del cambiamento intergenerazionale (Mannheim 2008; Inglehart, 1977) che offre una prospettiva utile per comprendere come lo scambio di conoscenze tra i giovani e gli anziani avviene in modo bidirezionale favorendo un flusso maggiore di conoscenze.

La ricerca di Park e Kim (2022) esamina l'uso quotidiano di smart speaker basati sull'intelligenza artificiale influenzano il benessere degli anziani soli con l'obiettivo di valutarne l'impatto nella vita quotidiana.

Il metodo utilizzato per raccogliere dati combina questionari e interviste, e lo studio suggerisce che l'integrazione di smart speaker basati su IA nella vita quotidiana degli anziani che vivono da soli può avere effetti positivi sul loro benessere. I risultati della ricerca indicano una riduzione della solitudine e un miglioramento nell'organizzazione delle attività quotidiane, pur evidenziando alcune criticità tecniche e di privacy. La ricerca ha coinvolto 291 partecipanti in Corea del Sud, con un'età media di circa 77 anni. I risultati hanno mostrato che l'uso frequente di questi dispositivi è associato a una riduzione della solitudine e della depressione tra gli anziani. In particolare, coloro che utilizzavano regolarmente gli smart speaker hanno riportato miglioramenti significativi nel loro stato emotivo e nella gestione delle attività quotidiane. Tuttavia, lo studio ha anche evidenziato alcune preoccupazioni riguardo alla privacy e alla sicurezza dei dati personali, sottolineando l'importanza di garantire la protezione delle informazioni sensibili degli utenti.

Inoltre, sul tema in oggetto per gli anziani ci sono preoccupazioni di tipo etico relative all'uso dell'IA poiché uno studio condotto da López e Molina (2021) ha dimostrato che il 65% degli intervistati è preoccupato per la privacy dei propri dati sanitari quando utilizza le tecnologie basate su IA.

Questi timori riflettono questioni più ampie legate alla governance dei dati, alla trasparenza degli algoritmi e al rischio di discriminazione o uso improprio delle informazioni.

Gli anziani, spesso meno familiari con le tecnologie digitali, possono sentirsi vulnerabili rispetto a potenziali violazioni della privacy o a una mancanza di controllo sui dati raccolti.

Per affrontare queste preoccupazioni, è fondamentale sviluppare sistemi di IA che siano non solo efficaci, ma anche eticamente progettati, con meccanismi chiari di consenso informato, protezione dei dati e responsabilità degli sviluppatori.

Promuovere la fiducia attraverso la trasparenza e l'educazione digitale degli utenti anziani è essenziale per favorire un'adozione più ampia e consapevole delle tecnologie assistive basate sull'IA.

Un ulteriore aspetto etico, meno spesso approfondito ma altrettanto cruciale, riguarda il possibile impoverimento delle relazioni umane.

Qui si inserisce il pensiero di Achille Ardigò (1988), che già decenni fa aveva previsto come la rivoluzione informatica potesse ridurre la qualità e la quantità dei rapporti interpersonali.

Ardigò sottolineava l'importanza del contatto umano autentico, mettendo in guardia contro il rischio che l'automazione e la tecnologia sostituissero, o quantomeno limitassero, le interazioni sociali fondamentali per il benessere emotivo e psicologico, soprattutto delle persone più vulnerabili come gli anziani. Questo richiamo è particolarmente pertinente nel contesto dell'IA in ambito sanitario, dove sebbene la tecnologia possa supportare e migliorare l'assistenza, non deve mai sostituire il rapporto umano, che rimane insostituibile per garantire il supporto emotivo e la dignità della persona.

2. Dalla telemedicina alla cura personalizzata: il ruolo dell'IA nell'assistenza agli anziani

Attraverso l'utilizzo della telemedicina e l'intelligenza artificiale viene offerta la possibilità di monitorare i pazienti e, nel caso degli anziani, questa tecnologia consente di tenere sotto controllo le condizioni croniche come l'ipertensione, il diabete o le malattie cardiovascolari riducendo la necessità di fare delle visite mediche (Chen, 2020).

Secondo alcuni studi (Bates *et al.,* 2019), la telemedicina ha dimostrato di migliorare l'accesso alle cure, ridurre i costi sanitari, gli esiti di salute dei pazienti anziani e offrire una gestione continua della salute, monitorando i parametri vitali come la pressione sanguigna, i livelli di glucosio e la frequenza cardiaca. Un'indagine condotta da Denecke *et al.* (2020) ha

rivelato che il 75% degli anziani che utilizzano dispositivi di monitoraggio remoto ha riportato un miglioramento nella gestione della salute, grazie alla possibilità di ricevere un riscontro immediato sui parametri vitali, il 65% di loro ha affermato di sentirsi più sicuro sapendo che i professionisti della salute monitorano costantemente le loro condizioni.

Inoltre, l'IA può integrarsi con dispositivi indossabili per rilevare eventuali cadute o segnali di emergenza, attivando automaticamente sistemi di supporto (Davenport, Kalakota, 2019).

L'automazione delle funzioni di monitoraggio e la gestione delle emergenze riduce il carico sui caregiver e sul personale sanitario, rendendo gli interventi più efficienti e tempestivi, con la necessità di garantire la privacy e la sicurezza dei dati sanitari.

Uno studio condotto da Robinson *et al.* (2013) ha evidenziato che l'interazione con robot sociali ha portato a un aumento del 30% della soddisfazione sociale tra gli anziani partecipanti mentre, il 40% degli utenti ha riferito di sentirsi meno soli grazie alla compagnia offerta da questi dispositivi. Alessa & Al-Khalifa (2023) hanno condotto uno studio su un sistema di conversazione basato su ChatGPT agli anziani, per dare compagnia e ridurre i sentimenti di solitudine e isolamento sociale.

I ricercatori hanno somministrato il test su un campione di utenti anziani per analizzare il grado di coinvolgimento e l'utilità percepita.

I risultati hanno mostrato che gli anziani trovano utile l'interazione con l'assistente virtuale, specialmente per combattere la solitudine ma, sono emerse alcune criticità, riguardanti la mancanza di personalizzazione delle risposte, il rischio di dipendenza dall'IA e questioni legate alla privacy.

Lo studio di Abdollahi *et al.* (2022) ha combinato l'intelligenza emotiva ed artificiale in un robot chiamato Ryan, per dare compagnia agli anziani con depressione e demenza. I risultati dello studio hanno mostrato un effetto positivo sull'umore degli utenti, suggerendo che i robot di assistenza possono essere strumenti efficaci per migliorare il benessere emotivo degli anziani. Il test è stato condotto su un gruppo di anziani, utilizzando strumenti di ricerca qualitativi e quantitativi per valutare l'impatto delle interazioni. Gli utenti coinvolti nell'indagine hanno riportato una maggiore sensazione di compagnia e un miglioramento del loro stato emotivo.

Questi studi dimostrano il potenziale dei robot nell'assistenza agli anziani, e sottolineano l'importanza di un'interazione più personalizzata e di un design che consideri le specifiche esigenze cognitive ed emotive di ciascun utente.

3. Tecnologie intelligenti per un'anzianità attiva e autonoma

L'intelligenza artificiale sta rivoluzionando la mobilità degli anziani attraverso dispositivi tecnologici avanzati che supportano l'autonomia e migliorano la qualità della vita degli anziani favorendo l'autosufficienza e la sicurezza.

La capacità di muoversi liberamente e svolgere attività quotidiane come fare la spesa, socializzare o prendersi cura della propria salute contribuisce notevolmente alla qualità della vita (Chen, 2020).

Tuttavia, con l'invecchiamento, diverse patologie, come l'artrite, le malattie neurologiche e le problematiche cardiovascolari, possono compromettere la mobilità, riducendo l'indipendenza e aumentando il rischio di isolamento sociale e depressione (Heinrich *et al.,* 2020).

Per Mendez et al. (2020), le protesi intelligenti permettono agli anziani di camminare e muoversi con maggiore facilità, riducendo il rischio di cadute e migliorando l'autosufficienza.

I benefici dell'IA riguardano in primo luogo, come l'IA consente una personalizzazione delle soluzioni, adattando le tecnologie alle esigenze individuali di ciascun anziano: ad esempio, i dispositivi intelligenti possono essere regolati in base ai cambiamenti nel livello di mobilità o nel comportamento del paziente, riducendo il rischio di incidenti o cadute e permettendo un rapido intervento in caso di emergenze.

L'adozione di tecnologie intelligenti riduce inoltre la necessità di assistenza fisica continua, consentendo agli anziani di rimanere a casa più a lungo e migliorando la loro qualità della vita.

Per Ardigò è fondamentale assicurarsi che l'uso dell'intelligenza artificiale non sostituisca completamente il contatto umano, poiché quest'ultimo rimane un elemento insostituibile per il benessere emotivo degli anziani.

Le tecnologie possono supportare e facilitare molte attività quotidiane, ma non possono replicare la profondità delle relazioni umane, l'empatia e il calore di un'interazione diretta. La presenza di familiari, amici o operatori sanitari è essenziale per prevenire sentimenti di isolamento e solitudine, problemi comuni tra gli anziani.

L'IA dovrebbe quindi essere vista come uno strumento complementare che integra, ma non rimpiazza, il supporto umano e solo mantenendo questo equilibrio sarà possibile promuovere un invecchiamento sano e soddisfacente. Infine, un aspetto ancora poco esplorato ma di crescente rilevanza riguarda la dimensione etica e sociale dell'adozione di tecnologie intelligenti nella popolazione anziana.

La formazione, l'accompagnamento all'uso e la progettazione inclusiva sono elementi fondamentali per assicurare che l'innovazione tecnologica contribuisca realmente a ridurre le disuguaglianze e a migliorare la qualità della vita di tutti gli anziani, indipendentemente dal loro sfondo socioeconomico. L'intelligenza artificiale e i dispositivi digitali devono essere accessibili a tutti, senza creare nuove forme di esclusione per chi ha meno risorse o meno familiarità con la tecnologia.

È importante che l'adozione di queste tecnologie non si trasformi in un ulteriore fattore di disuguaglianza sociale, lasciando indietro chi, per motivi economici, culturali o di alfabetizzazione digitale, non può beneficiarne. Per questo, insieme allo sviluppo tecnologico, devono andare di pari passo percorsi di formazione, supporto e accompagnamento, pensati per essere davvero inclusivi e comprensibili per tutti gli anziani.

Inoltre, è necessario che il rapporto tra tecnologia e persona anziana sia sempre fondato sul rispetto dell'autonomia e volontà di quest'ultima.

Le soluzioni tecnologiche non devono mai sostituire le scelte individuali o imporre modelli di comportamento, ma piuttosto affiancare e potenziare la capacità delle persone di vivere in modo dignitoso e consapevole.

In sintesi, l'etica nelle tecnologie per l'invecchiamento attivo deve sempre guidare lo sviluppo e l'applicazione di queste innovazioni, affinché siano strumenti di vera emancipazione e benessere, e non di esclusione o controllo.

Conclusioni

Come si è visto, la socializzazione è un elemento cruciale per il benessere psicologico e sociale degli anziani, e la tecnologia si configura come uno strumento fondamentale per facilitarla. Grazie a dispositivi digitali, social network e piattaforme di comunicazione, gli anziani possono mantenere legami significativi con familiari e amici, riducendo il rischio di isolamento ed emarginazione. Le teorie sociologiche e psicologiche, come quelle di Durkheim, Havighurst e Carstensen, dimostrano che l'integrazione sociale e l'attività mentale e fisica sono essenziali per un invecchiamento sano, e la tecnologia non solo permette agli anziani di restare attivi e informati, ma favorisce anche la selettività emotiva, consentendo loro di concentrarsi su relazioni significative e gratificanti.

L'IA può garantire un'inclusione sociale efficace e sostenibile poiché porta ad una trasformazione nei modelli di assistenza sanitaria, e le tecnologie come la telemedicina e l'intelligenza artificiale stanno emergendo come soluzioni innovative per migliorare la qualità della vita degli anziani.

La telemedicina consente un monitoraggio delle condizioni di salute, riducendo la necessità di visite fisiche e migliorando l'accesso alle cure, mentre l'IA offre strumenti avanzati per diagnosi più precise e dei trattamenti personalizzati.

Tuttavia, queste innovazioni presentano anche delle sfide, tra cui la necessità di garantire la sicurezza dei dati, l'accessibilità tecnologica per gli anziani e il mantenimento di un rapporto umano nell'assistenza sanitaria oltre che fare attenzione ad erogare delle cure personalizzate che tendono a rappresentare una grande opportunità per rendere il sistema sanitario più efficiente e sostenibile. È fondamentale che il progresso tecnologico sia accompagnato da un approccio umano e attento ai bisogni emotivi degli anziani poiché il divario digitale rappresenta uno degli ostacoli principali che devono affrontare nell'interazione con l'intelligenza artificiale.

Quindi, la difficoltà non è solo tecnologica, ma anche culturale, poiché il contesto in cui le diverse generazioni sono cresciute influisce sulla loro familiarità con le nuove tecnologie.

Tuttavia, la teoria del cambiamento intergenerazionale evidenzia come il dialogo tra giovani e anziani possa ridurre questo gap, favorendo uno scambio reciproco di conoscenze e valori, che permette ai giovani di arricchirsi grazie all'esperienza e alla saggezza delle generazioni più anziane. Nel promuovere questi scambi non solo si aiuta a colmare il divario digitale, ma si rafforzano anche i legami tra generazioni, creando una società più inclusiva e coesa.

In conclusione, come si evince dagli studi citati in precedenza, l'intelligenza artificiale, se integrata in modo etico e accessibile, può migliorare significativamente la qualità della vita degli anziani, garantendo loro una maggiore autonomia, sicurezza e inclusione sociale per promuovere l'interazione intergenerazionale.

Riferimenti bibliografici

Abdollahi H., Mahoor M.H., Zandie R., Siewierski J., Qualls S.H. (2022). Artificial emotional intelligence in socially assistive robots for older adults: A pilot study. *IEEE Transactions on Affective Computing*, 14(3): 2020-2032. https://doi.org/10.1109/TAFFC.2022.3143803

Ardigò A. (1988). *La sociologia della salute come nuova scienza sociale*. Bologna: il Mulino.

Baltes P.B., Reese H.W., Lipsitt L.P. (1980). Life-Span Developmental Psychology. *Annual Review of Psychology*, 31: 65-110.

Bates D.W., Cohen M., Leape L.L., et al. (2019). The impact of telemedicine on health outcomes and cost efficiency. *Journal of Medical Systems*, 43(5): 53-65.

Carstensen L.L. (1992). Social and emotional patterns in adulthood: Support for socioemotional selectivity theory. *Psychology and Aging*, 7(3): 331-338.

Chen L.K. (2020). Gerontechnology and artificial intelligence: Better care for older people. *Archives of Gerontology and Geriatrics*, 91: 104252.

Cornwell B., Waite L. (2009). Social disconnection and health in later life. *Journal of Health and Social Behavior*, 50(1): 31-48.

Havighurst R.J. (1963). Successful aging. *The Gerontologist*, 3(1): 8-13.

Davenport T., Kalakota R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, 6(2): 94-98.

Denecke K., May R., Borycki E.M., Kushniruk A. (2020). Evaluation of 1-Year in-home monitoring technology by home-dwelling older adults, family caregivers, and nurses. *Frontiers in Public Health*, 8: 518957. https://doi.org/10.3389/fpubh.2020.518957

Durkheim É. (1895). *Les règles de la méthode sociologique*. Paris: F. Alcan.

Havighurst R.J. (1963). Successful aging. In Williams R., Tibbits C., Donahue W. (eds.), *Process of aging*. New York: Antherton.

Heinrich S., Cieza A., Vogel T. (2020). Health and mobility in older adults: A global perspective. *Journal of Aging and Health*, 32(5): 653-669.

Inglehart R. (1977). *The Silent Revolution: Changing Values and Political Styles Among Western Publics*. Princeton (NJ): Princeton University Press.

ISTAT (2020). Invecchiamento attivo e condizioni di vita degli anziani in Italia. Dipartimento per le politiche della famiglia. https://www.famiglia.governo.it/it/politiche-e-attivita/analisi-e-valutazione/studi-e-ricerche-di-settore/invecchiamento-attivo/istat-2020-invecchiamento-attivo-e-condizioni-di-vita-degli-anziani-in-italia/

Lopez C.M., Molina M.J. (2021). Ethical considerations in the use of artificial intelligence for older adults. *Journal of Medical Ethics*, 47(9): 579-584. https://doi.org/10.1136/medethics-2021-107193

Mannheim K. (2008). *Le generazioni*. Bologna: il Mulino.

Mendez J., Santin O., Alonso M. (2020). Smart prosthetics and AI: Enhancing mobility in elderly populations. *Journal of Medical Technology*, 36(2): 103-112.

Park S., Kim B. (2022). The impact of everyday AI-based smart speaker use on the well-being of older adults living alone. *Technology in Society*, 71: 102133.

Robinson H., MacDonald B., Kerse N., Broadbent E. (2013). The psychosocial effects of a companion robot: A randomized controlled trial. *Journal of the American Medical Directors Association*, 14(9): 661-667.

WHO (2002). *A Policy Framework*. Geneva: World Health Organization.

Unmasking racial bias in medical AI: a narrative review of evidence and implications

by *Francesco Mastrocola*[*], *Elisabetta Ferrara*, *Oscar Genovesi*[**]

Artificial Intelligence (AI) is transforming healthcare, promising improvements in diagnosis, treatment, and patient outcomes. However, racial bias persists and feeds inequities. This review inspects how bias manifests in medical AI domains, identifying unrepresentative training data and proxy variables. It explores mitigation strategies and knowledge gaps, spotting the interdisciplinary approach to fortify equitable and accountable medical AI.

Keywords: AI; bias; clinical; data; healthcare; racial bias.

Smascherare i pregiudizi razziali nell'intelligenza artificiale medica: una revisione narrativa delle prove e delle implicazioni

L'intelligenza artificiale (IA) sta trasformando l'assistenza sanitaria, promettendo miglioramenti nella diagnosi, nel trattamento e negli esiti clinici dei pazienti. Tuttavia, i pregiudizi razziali persistono alimentando disuguaglianze. Questa revisione esamina come i pregiudizi si manifestino nell'IA medica, identificando dati di formazione non rappresentativi e variabili proxy. Esplora strategie di mitigazione e lacune di conoscenza, individuando l'approccio interdisciplinare per rafforzare un'IA medica equa e responsabile.

Parole chiave: IA; pregiudizio; clinico; dati; assistenza sanitaria; pregiudizio razziale.

Introduction

The integration of artificial intelligence (AI) in healthcare represents a revolution holding both promise and peril for health equity. While AI technologies offer unprecedented diagnostic capabilities, mounting evidence suggests these systems may strengthen racial disparities in healthcare delivery.

DOI: 10.5281/zenodo.17523874

[*] Università Telematica "Leonardo da Vinci". francesco.mastrocola@unidav.it, elisabetta.ferrara@unidav.it.

[**] Fondazione Università "G. d'Annunzio". genovesioscar@gmail.com.

Paragraphs "Introduction", 1, 2, and 3 are attributed to F. Mastrocola; paragraph 3.1, 3.2, 3.3 are attributed to E. Ferrara; paragraphs 4 and "Conclusions" are attributed to E. Ferrara and O. Genovesi. All authors reviewed the work, approving the final version.

Healthcare data reflects historical inequities in access, delivery, and documentation. Machine learning algorithms may learn and perpetuate discrimination patterns embedded within medical records (Parasuraman, Manzey, 2010). Automation bias exacerbates these challenges, as healthcare providers may exhibit excessive reliance on AI-generated recommendations, overlooking clinical information outside algorithmic parameters (Parikh, Teeple, Navathe, 2019; Gigerenzer, Hoffrage, Kleinbölting, 1991).
Clinical data sources, such as EHRs and diagnostic testing patterns, reflect historical utilization differences across racial groups, representing disparities in care access rather than biological variations (Montavon, Samek, Müller, 2018). AI systems trained on such data may encode systemic inequities, creating self-reinforcing cycles of discriminatory healthcare delivery. This narrative review examined mechanisms through which AI can foster racial bias in medicine, analyzing current approaches to bias detection and mitigation.

1. Intersectional Framework for Understanding AI Bias in Healthcare

Algorithmic discrimination operates through interconnected systems of oppression rather than isolated identity categories. Crenshaw's foundational work demonstrates that experiences at intersections of multiple identities cannot be understood as simply the sum of separate discriminations (Crenshaw, 1989). His influence helps to understand AI bias, as highlighted by recent discussions on 'Real Talk: Intersectionality and AI' (Howard, 2023). Current AI fairness approaches have critical limitations – even when language biases based on race, ethnicity, and gender are mitigated in word embedding models, biases persist against intersectional groups such as "Mexican American females" (Guo, Caliskan, 2021).
Building on Collins's matrix of domination theory, AI bias operates within existing power structures that simultaneously privilege and marginalize different groups (Collins, 2019). Contemporary examples include image recognition applications that identify gender particularly poorly for dark-skinned women (Buolamwini, Gebru, 2018).
Critical technology studies position technology as inherently political rather than neutral. Benjamin's "New Jim Code" argues that automation can hide, speed, and deepen discrimination while appearing benevolent (Benjamin, 2019). Noble's "Algorithms of Oppression" demonstrates how search engines reinforce racism (Noble, 2018), while Crawford's "Atlas of AI" reveals AI as "a technology of extraction" (Crawford, 2021). These systems embed

structural inequities from healthcare data and institutions into algorithmic processes.
The sociology of risk provides insights into how AI implementation creates new uncertainty and trust relationships. Brown and van Voorst's analysis reveals how AI technologies generate "cultures of hope" among professionals seeking technological solutions amid resource constraints, masking systematic disadvantages for intersectional populations (Brown, van Voorst, 2024). The "opacity" challenges of AI systems further complicate clinical understanding and decision-making (Burrell, 2016; Grote, Berens, 2020; Hawley, 2015).

2. Methodology

Literature was identified through PubMed, Scopus, and Web of Science databases using terms related to artificial intelligence, healthcare, and racial bias, covering publications through January 2025. We included peer-reviewed articles, theoretical frameworks, and policy analyses addressing algorithmic bias in healthcare contexts.

A narrative review approach was selected to synthesize diverse literature types and provide a comprehensive analysis across technical, ethical, and clinical perspectives.

3. Mechanisms of Algorithmic Bias in Healthcare

Racial bias in AI within healthcare operates through interrelated mechanisms. AI systems derive learning from historical medical records that embed documented patterns of healthcare inequities, creating differential misclassification bias (Gianfrancesco *et al.,* 2018). Algorithms trained on historically biased data systematically magnify existing healthcare disparities.

EHRs present three areas of bias amplification: missing data bias affecting marginalized populations with fragmented care, sample size bias when insufficient minority data leads to algorithmic defaulting to majority trends, and measurement error bias from suboptimal care patterns (Burrell, 2016). These biases manifest through "automation complacency", where healthcare providers demonstrate excessive reliance on algorithmic recommendations while overlooking clinical information outside algorithmic parameters (Topol, 2019).

Mitigation strategies include preprocessing approaches such as weighting methods and data augmentation for underrepresented groups (Cary *et al.*, 2023). However, debate persists regarding race and ethnicity variables in clinical algorithms (Norgeot, Glicksberg, Butte, 2019; Cross, Choma, Onofrey, 2024), reflecting the challenge of balancing demographically aware algorithms against perpetuating societal biases.

3.1. Evolution and Current Challenges of AI Prediction Models in Healthcare

Healthcare prediction models have evolved from scoring systems with small datasets to advanced AI algorithms analyzing complex, multimodal data. While these systems show promise in diagnostic accuracy and personalized medicine, mounting evidence reveals racial bias concerns (Gianfrancesco *et al.*, 2018). Obermeyer *et al.* (2019) demonstrated how a widely used healthcare algorithm exhibited racial bias, reducing identification of Black patients needing additional care by more than half compared to White patients.

Such biases occur through unrepresentative training data, measurement classification bias, and encoding of historical healthcare disparities into algorithmic systems. Data quality issues compound these problems, including missing data disproportionately affecting marginalized populations (Nijman *et al.*, 2022) and measurement errors from medical devices performing differently across racial groups (Charpignon *et al.*, 2023). The WHO (2021) emphasized that AI's promise for improving healthcare worldwide can only be achieved by placing ethics and human rights at the center of design, deployment, and use.

3.2. Clinical Implementation and Healthcare Provider Perspectives

AI tools operate within environments where clinical judgment, situated cognition, and systemic biases converge. Chowdhury and Lake (2018) distinguish between explainability (mathematical aspects) and understandability (interpreting AI recommendations), crucial for recognizing bias perpetuation. The impact is evident in risk-stratification algorithms – Vyas, Eisenstein, and Jones (2020) found that race-adjusted algorithms in cardiac surgery could steer Black patients away from life-saving procedures based on poorly understood racial adjustments. The addition of AI risks can repropose a paternalistic model

of medical decision-making, where the 'computer knows best,' potentially undermining shared decision-making inside the inter-relations between clinicians and patients (McDougall, 2019).

Van de Sande *et al.* (2022) identify four key intervention domains: healthcare provider education in bias recognition, standardized data quality protocols, evolving regulatory frameworks with mandatory equity audits, and meaningful stakeholder engagement throughout AI development (Dankwa-Mullan *et al*., 2021).

3.2.1. Global Perspectives: AI Bias Challenges in Low and Middle-Income Countries

While much literature on AI bias emerges from high-income countries, unique challenges appear in low – and middle – income countries (LMICs), where AI systems may exacerbate rather than address health disparities. Critical "contextual bias" emerges when AI models trained on high-income country data are deployed in LMIC settings without adequate validation (Alami *et al.*, 2020). More than half of clinical AI datasets originate from the US or China, with almost all top databases affiliated with high-income countries (Larrazabal *et al*., 2021).

Studies across Latin America, Sub-Saharan Africa, and South Asia reveal that AI models trained on high-income country data may introduce substantial bias, leading to poor performance, particularly harmful in resource-limited settings (Schwalbe, Wahl, 2020). Dermatology AI systems trained predominantly on lighter skin tones showed reduced accuracy for darker skin conditions prevalent in African populations (Kamulegeya *et al*., 2019).

Limited local training data creates "data poverty" feedback loops where populations in data-rich regions benefit substantially more from AI healthcare applications, entrenching global health disparities. Cultural and linguistic factors further complicate implementation, as health concepts, symptom descriptions, and care-seeking behaviors vary across contexts, yet most AI systems are developed with limited consideration of these variations.

3.3. Implementation Strategies and Quality Assurance Frameworks

The regulatory frame for AI in healthcare is evolving, with animated debates concerning its classification as a medical device and the associated ethical and legal implications (Pesapane *et al.*, 2018; Gerke, Minssen, Cohen, 2020).

Addressing clinical implementation and global context challenges requires structured frameworks guiding AI system development, validation, and deployment while maintaining an equity focus. Van de Sande *et al*. (2022) propose step-by-step AI development with bias detection at each stage, aligning with the "roadmap for responsible machine learning in healthcare, "emphasizing data preprocessing, model development, and validation stages to prevent harm to vulnerable populations (Wiens *et al*., 2019).

Quality assurance frameworks serve as core implementation tools. The PROBAST tool (Wolf *et al*., 2019) offers structured bias risk assessment in prediction model studies, while the CONSORT-AI extension (Liu *et al*., 2020) provides clinical trial reporting guidelines for AI interventions. These frameworks emphasize transparent reporting of model development processes and thorough bias assessment across population subgroups.

Bellamy et al. (2019) describe the AI Fairness 360 toolkit, offering an extensible framework for detecting and mitigating algorithmic bias through preprocessing techniques for data debiasing and post-processing methods adjusting model outputs for demographic group fairness. The WHO's AI ethics and governance guidance (2021) recommends standardized bias detection approaches, including regular equity audits, diverse development team representation, continuous demographic performance monitoring, and transparent limitation reporting.

Validation processes must address disparities in model performance across racial and ethnic groups (Navarro *et al*., 2021). Future investigations should combine predominantly European and Anglo-Saxon studies with LMIC research to avoid Western-centric perspectives limiting generalizability and embrace approaches that prevent geographical, socio-economic, and cultural biases.

4. Limitations

Study limitations include the narrative review methodology, which may introduce selection bias in literature inclusion. The predominantly Western-centric literature limits generalizability to global contexts, and the rapid evolution

of AI technology may outpace current findings. Success requires ongoing collaboration between technology developers, healthcare providers, policy makers, and affected communities.

Conclusions

This review examined racial bias in medical AI, revealing that technical sophistication does not inherently protect against bias. Effective mitigation requires intervention at multiple levels, from data collection to clinical implementation. Key findings indicate AI systems risk amplifying existing disparities through algorithmic processes. Future research should focus on developing standardized bias detection methods and validation frameworks for diverse populations.

References

Alami H., Lehoux P., Auclair Y., de Guise M., Gagnon M.P., Fortin J.P., et al. (2020). Artificial intelligence in health care: laying the foundation for responsible, sustainable, and inclusive innovation in low- and middle-income countries. *Globalization and Health*, 19: 52. https://doi.org/10.1186/s12992-020-00584-1

Bellamy R.K.E., Dey K., Hind M., Hoffman S.C., Houde S., Kannan K., Lohia P., Martino J., Mehta S., Mojsilović A., Nagar S., Natesan Ramamurthy K., Richards J., Saha D., Sattigeri P., Singh M., Varshney K.R., Zhang Y. (2019). AI Fairness 360: an extensible toolkit for detecting and mitigating algorithmic bias. *IBM Journal of Research & Development*, 63(4/5): 1-15. https://doi.org/10.1147/JRD.2019.2942287

Benjamin R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge: Polity Press.

Brown P., van Voorst R. (2024). The influence of artificial intelligence within health-related risk work: a critical framework and lines of empirical inquiry. *Health, Risk & Society*, 26(7-8): 301-316. https://doi.org/10.1080/13698575.2024.2412374

Buolamwini J., Gebru T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81: 77-91.

Burrell J. (2016). How the machine 'thinks': understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1): 1-12. https://doi.org/10.1177/2053951715622512

Cary M.P. Jr., Zink A., Wei S., Olson A., Yan M., Senior R., Bessias S., Gadhoumi K., Jean-Pierre G., Wang D., Ledbetter L.S., Economou-Zavlanos N.J., Obermeyer Z., Pencina M.J. (2023). Mitigating racial and ethnic bias and advancing health equity in clinical algorithms: a scoping review. *Health Affairs (Millwood)*, 42(10): 1359-1368. https://doi.org/10.1377/hlthaff.2023.00553

Charpignon M.L., Byers J., Cabral S., Celi L.A., Fernandes F., Gallifant J., Lough M.E., Mlombwa D., Moukheiber L., Ong B.A., Panitchote A., William W., Wong A.I., Nazer L. (2023). Critical bias in critical care devices. *Critical Care Clinics*, 39(4): 795-813. https://doi.org/10.1016/j.ccc.2023.02.005

Chowdhury R., Lake M. (2018). Is explainability enough? Why we need understandable AI. *Forbes*. https://www.forbes.com/sites/rummanchowdhury/2018/06/04/is-explainability-enough-why-we-need-understandable-ai/?sh=33ed372d62f4

Collins P.H. (2019). *Intersectionality as Critical Social Theory*. Durham: Duke University Press.

Crawford K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press.

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.

Cross J.L., Choma M.A., Onofrey J.A. (2024). Bias in medical AI: implications for clinical decision-making. *PLOS Digital Health*, 3(11): e0000651. https://doi.org/10.1371/journal.pdig.0000651

Dankwa-Mullan I., Scheufele E.L., Matheny M., Quintana Y., Chapman W., Jackson G., South B.R. (2021). A proposed framework on integrating health equity and racial justice into the artificial intelligence development lifecycle. *Journal of Health Care for the Poor and Underserved*, 32(2): 300-317. https://doi.org/10.1353/hpu.2021.0065

Gerke S., Minssen T., Cohen G. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare. In *Artificial Intelligence in Healthcare*, 295-336. https://doi.org/10.1016/B978-0-12-818438-7.00012-5

Gianfrancesco M.A., Tamang S., Yazdany J., Schmajuk G. (2018). Potential biases in machine learning algorithms using electronic health record data. *JAMA Internal Medicine*, 178(11): 1544-1547. https://doi.org/10.1001/jamainternmed.2018.3763

Gigerenzer G., Hoffrage U., Kleinbölting H. (1991). Probabilistic mental models: a Brunswikian theory of confidence. *Psychological Review*, 98(4): 506-528. https://doi.org/10.1037/0033-295X.98.4.506

Grote T., Berens P. (2020). On the ethics of algorithmic decision-making in healthcare. *Journal of Medical Ethics*, 46(3): 205-211. https://doi.org/10.1136/medethics-2019-105586

Guo W., Caliskan A. (2021). Detecting emergent intersectional biases: contextualized word embeddings contain a distribution of human-like biases. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*: 122-133.

Hawley K. (2015). Trust and distrust between patient and doctor. *Journal of Evaluation in Clinical Practice*, 21(5): 798-801. https://doi.org/10.1111/jep.12374

Howard A. (2023). Real talk: intersectionality and AI. *MIT Sloan Management Review*. https://sloanreview.mit.edu/article/real-talk-intersectionality-and-ai/

Kamulegeya L.H., Okello M., Bwanika J.M., Musinguzi D., Nakibuuka J., Bassajja A., et al. (2019). Using artificial intelligence on dermatology conditions in Uganda: a case for diversity in training data sets for machine learning. *bioRxiv*. https://doi.org/10.1101/826057

Larrazabal A.J., Nieto N., Peterson V., Milone D.H., Ferrante E. (2021). Sources of bias in artificial intelligence that perpetuate healthcare disparities – a global review. *PLOS Digital Health*, 1(1): e0000022. https://doi.org/10.1371/journal.pdig.0000022

Liu X., Cruz Rivera S., Moher D., Calvert M.J., Denniston A.K., SPIRIT-AI and CONSORT-AI Working Group (2020). Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nature Medicine*, 26: 1364-1374. https://doi.org/10.1038/s41591-020-1034-x

McDougall R.J. (2019). Computer knows best? The need for value-flexibility in medical AI. *Journal of Medical Ethics*, 45(3): 156-160. https://doi.org/10.1136/medethics-2018-105118

Montavon G., Samek W., Müller K.R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73: 1-15. https://doi.org/10.1016/j.dsp.2017.10.011

Navarro C.L.A., Damen J.A.A., Takada T., Nijman S.W.J., Dhiman P., Ma J., Collins G.S., Bajpai R., Riley R.D., Moons K.G.M., Hooft L. (2021). Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. *BMJ*, 375(2281): 1-9. https://doi.org/10.1136/bmj.n2281

Nijman S., Leeuwenberg A.M., Beekers I., Verkouter I., Jacobs J., Bots M.L., Asselbergs F.W., Moons K., Debray T. (2022). Missing data is poorly handled and reported in prediction model studies using machine learning: a literature review. *Journal of Clinical Epidemiology*, 142: 218-229. https://doi.org/10.1016/j.jclinepi.2021.11.023

Noble S.U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.

Norgeot B., Glicksberg B.S., Butte A.J. (2019). A call for deep-learning healthcare. *Nature Medicine*, 25(1): 14-15. https://doi.org/10.1038/s41591-018-0320-3

Obermeyer Z., Powers B., Vogeli C., Mullainathan S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464): 447-453. https://doi.org/10.1126/science.aax2342

Parasuraman R., Manzey D.H. (2010). Complacency and bias in human use of automation: an attentional integration. *Human Factors*, 52(3): 381-410. https://doi.org/10.1177/0018720810376055

Parikh R.B., Teeple S., Navathe A.S. (2019). Addressing bias in artificial intelligence in health care. *JAMA*, 322(24): 2377-2378. https://doi.org/10.1001/jama.2019.18058

Pesapane F., Volonté C., Codari M., Sardanelli F. (2018). Artificial intelligence as a medical device in radiology: ethical and regulatory issues in Europe and the United States. *Insights into Imaging*, 9(5): 745-753. https://doi.org/10.1007/s13244-018-0645-y

Schwalbe N., Wahl B. (2020). Artificial intelligence and the future of global health. *The Lancet*, 395(10236): 1579-1586. https://doi.org/10.1016/S0140-6736(20)30226-9

Topol E.J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1): 44-56. https://doi.org/10.1038/s41591-018-0300-7

Van de Sande D., Van Genderen M.E., Smit J.M., Huiskens J., Visser J.J., Veen R.E.R., van Unen E., Ba O.H., Gommers D., Bommel J.V. (2022). Developing, implementing, and governing artificial intelligence in medicine: a step-by-step approach to prevent an artificial intelligence winter. *BMJ Health & Care Informatics*, 29(1): e100495. https://doi.org/10.1136/bmjhci-2021-100495

Vyas D.A., Eisenstein L.G., Jones D.S. (2020). Hidden in plain sight — reconsidering the use of race correction in clinical algorithms. *New England Journal of Medicine*, 383(9): 874-882. https://doi.org/10.1056/NEJMms2004740

Wiens J., Saria S., Sendak M., Ghassemi M., Liu V.X., Doshi-Velez F., Jung K., Heller K., Kale D., Saeed M., Ossorio P.N., Thadaney-Israni S., Goldenberg A. (2019). Do no harm: a roadmap for responsible machine learning for health care. *Nature Medicine*, 25(9): 1337-1340. https://doi.org/10.1038/s41591-019-0548-6

Wolf R.F., Moons K.G.M., Riley R.D., Whiting P.F., Westwood M., Collins G.S., Reitsma J.B., Kleijnen J., Mallett S., PROBAST Group (2019). PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine*, 170(1): 51-58. https://doi.org/10.7326/M18-1376

World Health Organization (2021). *Ethics and governance of artificial intelligence for health: WHO guidance*. Geneva: World Health Organization. https://apps.who.int/iris/bitstream/handle/10665/341996/9789240029200-eng.pdf

Transform or combine? Tracing the irreducibility of human actions with respect to chatbot by examining definitions and evaluation tests

by *Simone D'Alessandro*[*]

Are creativity and intelligence distinct or coinciding concepts? How do they determine distinctions between humans and chatbots? Is it also possible to distinguish between incapacities? Integrating ethnomethodology and discursive analysis, the research examines the differences between human and A.I. incapacities by analysing the theoretical assumptions and international tests used by programmers to evaluate interactions: a) Turing Test; b) Winograd and Winogrande Test; c) Lovelace Test.

*Keyword*s: automatism; chatbot; creativity; incapacity; artificial intelligence; semantics.

Trasformare o combinare? Tracciare l'irriducibilità delle azioni umane rispetto ai chatbot esaminando definizioni e test di valutazione

Creatività e intelligenza sono concetti distinti o coincidenti? Come determinano le distinzioni tra esseri umani e chatbot? È possibile distinguere anche le incapacità? Integrando etnometodologia e analisi discorsiva, la ricerca esamina le differenze tra le incapacità umane e quelle delle I.A. analizzando i presupposti teorici e i test internazionali utilizzati dai programmatori per valutare le interazioni: a) Test di Turing; b) Test di Winograd e Winogrande; c) Test di Lovelace.

Parole chiave: automatismo; chatbot; creatività; incapacità; intelligenza artificiale; semantica.

Introduction: the semantic ambiguities of assumptions and research paths linking creativity and intelligence

Ethnomethodology has shown how banal or common-sense social discourses, actions, and phenomena reveal more revolutionary, creative, and counter-intuitive significance than we expect at first naive observation.

DOI: 10.5281/zenodo.17522282

[*] Università degli Studi G. D'Annunzio di Chieti-Pescara. simone.dalessandro@unich.it.

It all depends on what questions we ask, how our attention is captured, and what starting assumptions we accept, taking for granted definitions that we do not subject to further falsification (Garfinkel, 1991).

Indeed, it would always be appropriate to subject the assumptions of definitions to careful investigation in order to understand whether they favour self-deception (Zimmerman, Pollner, 1983). When researchers circumscribe definitions of words that imply complex concepts, they construct a concatenation of associations with other terms that they consider compatible with their own discursive repertoire.

Such repertoires imply cultural and ideological assumptions that constrain explicit words, generating specific constructions of meaning.

This also happens with the terms like intelligence and creativity. Definitions emerge that at first seem clear, but in the process of association with other terms, present ambiguities that scholars tend to disambiguate, selecting a reductionist path that makes the research objectifiable.

From the definitions given to intelligence, complementary relationships emerge with the term 'creativity' and its repertoires, where the ability to adapt to new situations and the capacity to transform reality in a useful and improving way are emphasised.

I will dwell specifically on the ambivalent relationships between intelligence and creativity, because from these derive distinct ways of understanding artificial intelligence, particularly that emerging from the *machine-learning* algorithms used in the large language models of chatbots (LLM). From these distinctions emerge, in turn, different ways of understanding concepts apparently opposed to the concept of intelligence: human and artificial incapability. By human inability, I mean the subjective limits of understanding of what is said or shown by a human being: in this sense, each subject has its own deficient nature.

By artificial inability I mean an objective set of limitations: 1. Inability to understand semantics; 2. Inability to contextualize conversational assumptions and implicatures; 3. Inability to interact or act unpredictably; 4. Inability to decide in the absence of starting information provided by the program; 5. Inability to boycott the automation of programming systems. Should creativity and intelligence be thought of as two distinct or coinciding aspects? The answer to this research question depends on the definitions selected by researchers. Theories always depend on social constructions and cultural ways of understanding discourse terms semantically. Anastasi and Schaefer (1971) state that creativity and intelligence are not clearly distinguishable concepts. De Caroli examines studies that support an interdependent relationship between Creativity and IQ, but also studies that

deny this correlation (De Caroli, 2016). Getzels' (1962) empirical research shows that there is a low correlation between creativity and intelligence.

Arieti (1990) and Klein (2022) confirm Getzels' thesis. Polanyi (1979), Sternberg (1988) and Sennett (2008) argue that creative processes depend on tacit knowledge, manipulation of reality and serendipity.

Hadamard (2022) emphasises the relationship between emotion and cognition. Baron-Cohen argues that autism drives specific creative and inventive processes (Baron-Cohen, 2021). Power establishes a link between creativity, schizophrenia, and bipolar disorder (Power, 2015). There are theories that distinguish intelligence from creativity and theories that start from inclusive definitions.

Researchers who start from inclusive definitions believe that creative behaviour is made up of a mix of capabilities and incapabilities. A creative subject manifests an intelligence capable of overturning the assumptions of a discourse, but incapable (or poorly capable) of being analytical.

The creative person connects distant semantic concepts (through metaphors) but is less able to connect logically close concepts.

Can we then say that, under certain conditions, incapacity can be considered a form of intelligence useful for experimenting with the creative process?

Bergson, in his work *Évolution créatrice* (Bergson, 1907; tr. it. 2012) defines automatic behaviour as 'instinct possessing its own reasons'.

Taking up this argument, social scientists have advanced the following question: what relationship does critical intelligence establish with the automatisms of routines? Is there a continuity between automatic responses and those that rebel against automatism?

There is a close correlation between the way the question is posed and the way the research path is constructed. In this essay, I will attempt to understand the relationships between combinatorial and transformative creative process by answering the following interrelated research questions: 1. Are creativity and intelligence distinct or coinciding concepts? 2. How do they determine distinctions between humans and chatbots? 3. Is it also possible to distinguish between incapacities?

I will attempt to infer clear distinctions between human and chatbot, including elements that have traditionally been discarded by researchers: ambivalences in the relationship between intelligence, creativity, inability, and automatism. In order to answer the questions posed, I will examine the tests used by programmers to distinguish human from non-human behaviour: a) the classic Turing test; b) the Winograd test and the Winogrande test; c) the Lovelace test on the creativity of artificial agents.

1. The Turing Test and the simulation of interactions: human versus artificial automatism

The Turing Test is a method for testing whether a machine, speaking through a computer interface, can be mistaken for a human being. In the test, the syntactic manifestation of verbal and para-verbal signals, emerging in the course of an interaction in the form of a dialogue (written or oral), enables the machine to simulate human thoughts, conversations and (indirectly) behaviour. The classic test was based on the relationship between three participants. Let us assume a dialogue between A and B where the two interlocutors do not see each other or cannot verify each other's identity.

We insert a third interlocutor C (human questioner). We also add that A must help C, while B (non-human) must deceive him. If C cannot distinguish A's behaviour from B's behaviour, why not attribute intelligence to B as well? (Turing, 1950). On the level of formal rules, Turing's argument is logically founded. On the other hand, if we analyse this assumption in sociological terms, the intention is evidently reductionist, as it simplifies a relationship by circumscribing it within a 'simulation of interaction', eliminating the ambivalent signals (intentional and unintentional) of non-verbal communication such as facial expressions, postures and involuntary body movements. Moreover, the test limits para-verbal expressions to the use of punctuation (in the case of textual chatbots) or to non-expressive sound intonations (in the case of vocal chatbots).

Another weakness of the Turing test, in terms of agency, is its reliance on deception as chatbots that successfully pass the Turing Test – which has been updated repeatedly to date – manage to fool humans for short periods of time, partly by evading questions (Riedl, 2014). But the crux of the debate is on how to define and, consequently, conceive the concept of comprehension.

According to a reductionist view, 'comprehension' of language simply means knowing how to use it.

According to an anti-reductionist understanding it means a) being able to grasp the nuances and latent assumptions of each sentence uttered; b) being able to contextualise the meaning of words and phrases according to the specific circumstances of a given situation and with respect to the values and cultural background of the interlocutors (Grice, 1996).

The basic issue is not related to the empirical administration of the test, but to the assumptions we make on the definitional level when we decide to give certain attributes to the terms under consideration and the resulting relationship with the alter. As the sociological literature that has dealt with

the topic shows, judging the degree of intelligence of certain behaviours depends on how we, as human beings, choose the criteria that enable the social construction of specific value judgements (Mazzotti, 2015).

If we look at the Turing test under a sociological lens, we could reverse the purpose of the test itself, stating that this method does not allow us to understand whether a machine is intelligent because it can deceive a human observer by simulating credible interactions. Instead, the test would show the extent to which humans can have automatisms that direct and influence the expectations of our conversations. Reversing Turing's perspective, we could say that the test allows us to better explain how certain habitual patterns of human beings function. Pareto (1916) had already shown the extent to which human beings act on the basis of repeated combinations.

Intelligence and creativity are related to habits and automatisms that we could also call 'action programmes' that allow us to enter into relationships without necessary self-reflection. After all, most human conversations are stereotyped and follow predictable and mechanical flows.

Think of introductory phrases, clichés, rituals, conversational turns: elements of interaction examined by the Palo Alto school, symbolic interactionism, and ethnomethodology.

The term 'automaton', which has been used since the third century BC to represent objects and artefacts that simulate the behaviour of the living, derives from a Greek term with an ambivalent meaning: *αὐτόματος*.

This term denotes an (s)object 'acting of its own will', but in common usage it connotes 'behaving like an automaton in the sense of acting mechanically and without thinking'.

Consequently, «we should not look for the automaton among machines, as we would naturally do (...) Rather, following a Bergsonian indication, its archetype should be sought among living beings» (Ronchi, 2021: 2).

Can the recurring, predictable and automatic aspects of the human (easily reproduced by an artificial intelligence) be defined as fully intelligent? If these automatisms produce something unexpected, can they be considered creative? Programs capable of generating dialogues show that there is a creative part that is recursive, i.e.: based on a few rules that can endlessly generate new productions that can have meaning and originality for the human being who observes and interprets them. In this case we can speak of repetitive and fractal meta-rules of creativity (D'Alessandro, 2023): a part that is present within an algorithm but is also possessed by the human being in terms of routines. This part is what unites the human with the artificial non-human, and we could call it combinatorial creativity (Boden, 1990). Creativity that applies rules that can create other rules to change and generate

something new in terms of a) assembling elements; b) subtraction-partition; c) reversal of data (D'Alessandro, 2025). However, while combinatorial creativity recombines prior knowledge, transformative creativity invents new categories that enable the emergence of the unexpected (Klein, 2022).

This type of transformative intelligence is related to the human capacity to interpret what is happening in a different way, reversing the perspective, or cancelling it out altogether: the person voluntarily decides to unpredictably distort an automatic path, generating paradoxes that allow established knowledge to be transcended, creating other content.

If the machine-learning algorithm generates on the basis of existing data combinations, the human being is able to 'renounce' the exploration of a topic by constructing other meanings.

The Turing test does not prove that machines can understand the non-automatic and unpredictable part of the human.

2. Winograd and Winogrande: the differences between human and artificial stupidity

Terry Winograd invented a test (still used today in his later reworkings) that demonstrates the inability of artificial intelligence, i.e., its specific inability that is very different from that of humans (Winograd, 1972).

The test is characterised by the administration of ambiguous sentences.

Winograd's test has several versions, called schemes, consisting of two sentences that differ only by one or two words, but contain an ambiguity that is resolved in opposite ways.

It is not possible to pass the test using only syntax rules. There is a need for semantic understanding and contextualisation of reality.

This is an example of Winograd's scheme: a) The city councillors refused permission to the demonstrators because they feared riots; b) The city councillors refused permission to the demonstrators because they instigated riots. People interpret the first sentence to mean that it is the city councillors who fear riots; they interpret the second sentence to mean that the instigators are the protesters. The sentences are structurally identical, but the human being, by contextualising the meaning of the two sentences and 'understanding' the roles and tasks of the city councillors and the protesters, selects the two distinct subjects. In conversations and relationships between human beings, sentences like these are frequent. When not properly understood, they give rise to misunderstandings and trigger conflicts. However, in the course of human conversation, repeated contextualisation

attempts allow misunderstandings to be resolved. Conflict may persist due to attitudes that are beyond rational comprehension (stubbornness, dislike, amusement, etc.) or that are motivated by other rational calculations (e.g., a definite intention to build controversy in order to break off relations with an interlocutor). Artificial intelligence, on the contrary, fails to contextualise and fails in an act that human beings consider simple.

There are 150 examples of Winograd schemes blocking chat bots. Because chat bots do not understand conversational implicatures[1]. Implicatures are inferences that an interlocutor makes when talking to another interlocutor, trying to understand the implicit dimensions that depend not only on the sentence, but also on the intentions and expectations of the other. They can be conventional (dependent on the shared meaning of the words used) or conversational. Conversational implicatures depend on the cooperative contribution and accepted orientations in the conversation (Grice, 1996).

In 2012, a group of researchers at New York University perfected Winograd's test by using pairs of sentences that differ by only one word (containing an object-complement pronoun that reverses the meaning of the sentence), which are followed by two questions, one for each sentence, e.g.: 1) I poured milk from the container into the jug until it was full; question: what is full, the container or the jug? 2) I poured the milk from the container into the jug until it was empty; question: what is empty, the container or the jug? With the improved Winograd, researchers confirm the semantic inability of chat boxes.

In 2019, a second group of researchers from the Allen Institute for Artificial Intelligence created 'Winogrande': a test with 44,000 sentences on a variety of topics. "While humans scored very high, the language models of the neural network (...) scored much lower"[2].

3. The Lovelace Test and exploratory creativity

The Lovelace 2.0. is a test that seeks to measure the creative capacity of a computational system, attempting to formalise the notions of originality and surprise (Bringsjord, Bello, Ferrucci, 2001).

[1] You can find the collection of such examples at:
https://cs.nyu.edu/~davise/papers/WinogradSchemas/WSCollection.html
[2] Sakaguchi K., Le Bras R., Bhagavatula C., Choi Y. (2019).

Riedl (2014) proposes an articulated test to show that a certain subset of creative acts exclusively requires human intelligence. The test is called Lovelace 2.0. Here the artificial agent is challenged on the basis of the following rules of engagement: 1. 'a' (artificial agent) must create an artefact 'o' of type 't' (output as the of processes that can be repeated and are not random hardware errors); 2. 'o' must conform to a set of constraints 'C' where ci ∈ C and is any criterion expressible also in natural language; 3. A human evaluator 'h', having chosen t and C, is satisfied that 'o' is a valid variation of t and satisfies C; 4. a human arbiter determines that the combination of t and C is not unrealistic for a human.

The constraints set make the test Google-proof and resistant to Searle's Chinese Room arguments[3]. An evaluator may impose the constraints he deems necessary to ensure that the system produces a surprising artefact or story with an original subject. Although C need not be expressed in natural language, the set of possible constraints must be equivalent to the set of all concepts that can be expressed by a human mind.

The Lovelace 2.0 Test is designed to encourage scepticism in human evaluators. This test assumes that creativity is a distinctive trait of human intelligence, but not an exclusive one.

Once again, the type of test is influenced by the starting assumptions made by the researcher. So far, human evaluators have not been surprised by machine responses. A.I.'s creativity is confirmed combinatory but does not show unpredictable and transformative behaviour.

Conclusions

As I have tried to show in my research work, the irreducibility between human intelligence and chatbot depends on the theoretical and discursive assumptions accepted by researchers investigating these fields.

I have highlighted the irreducible differences between human intelligence, creativity, automatisms and incapacities with respect to artificial expert systems, analysing theoretical assumptions and tests used by

[3] In the present paper, I have not examined Searle's philosophical experiment of the Chinese Room, presented in the article Minds, Brains and Programs, published in 1980 in the scientific journal The Behavioural and Brain Sciences, because it has not been turned into a test that can be submitted to a chat bot:
https://plato.stanford.edu/archives/win2020/entries/chinese-room/

programmers to distinguish human conversation from non-human interactions, in particular: a) The Classical Turing Test; b) The Winograd Test and The Winogrande Test; c) The Lovelace 2.0 Test.

Going back to the questions posed in the introduction, I highlight the most important factors that I understood during the research phases:

1. Are creativity and intelligence distinct or coinciding concepts? Researchers today are divided on this issue. There are theories that exclude correlations between intelligence and creativity; theories that are inclusive; theories that are ambivalent. Each starting definition generates different relationships. In some cases, these definitions have generated evaluation tests. The Turing Test assumes intelligence as the ability to emulate and dissimulate behaviour through verbal and paraverbal language, excluding the concept of comprehension and non-verbal forms of communication. The Winograd and Winogrande tests probe the semantic inability of the machine through sentences that can only be understood if they are contextualised. The Lovelace 2.0 test measures the machine's inability to 'surprise' the human evaluator. Each of these tests arises on the basis of assumptions and implications hidden in the definitions given to the terms intelligence and creativity. The way of selecting the definition determined the way of testing A.I.
2. Do Creativity and Intelligence determine irreducible distinctions between artificial human and non-human? The tests adopted by the programmers showed the presence of A.I. incapacities completely different from human ones: 1. Inability to understand semantics; 2. Inability to contextualise conversational implicatures; 3. Inability to respond or act unpredictably; 4. Inability to decide in the absence of starting information; 5. Inability to boycott the automatism of programming rules.
3. Can creativity and intelligence also imply automatism and incapacity as additional heuristic resources? Again, this all depends on how concepts are defined and what assumptions are involved in the definitions. Human beings often act automatically, demonstrating a will that would appear to be devoid of thought and equipped with habitual rules. Yet, the human's automatic mode of being is different from the artificial one in that it can be interrupted by consciousness. Moreover, the scientific community has not agreed on the concept of automatism: can what is automatic in the human and, consequently, reproducible by an artificial intelligence, be defined as intelligent? Among researchers who make distinctions

between human and artificial, the point is not technological, but cultural and ideological. The Lovelace test shows that AI is endowed with combinatorial creativity, but not transformative.

In conclusion, we can state that theoretical assumptions and tests prove that humans are more mechanical than they think they are. However, tests do not prove that machines can understand the non-mechanical part of the human. AI represents a distinct form of agency with programmable goals, data-driven adaptability, and distributed functionality. Unlike human agents, AI lacks consciousness, intentionality, and intelligence (Floridi, 2025).

The conclusions of the article can also be summarized in the following table:

Test	**Purpose**	**AI Capabilities Tested**	**Human vs AI Distinction**	**Creativity Type Involved**	**Main Limitations of AI**
Turing	Assess whether a machine can imitate human conversation	Emulation of human-like responses	Focuses on surface-level imitation, not comprehension	Combinatorial (rule-based recombination)	Lacks semantic understanding Cannot interpret non-verbal cues Relies on deception
Winograd / Winogrande	Test semantic and contextual understanding	Contextual disambiguation and implicature comprehension	Humans resolve ambiguity through world knowledge; AI fails without explicit cues	None (focus is on comprehension, not creation)	Inability to resolve ambiguity Fails to grasp conversational implicatures
Lovelace 2.0	Evaluate AI's creative capacity under constraints	Originality, surprise, and constraint satisfaction	AI shows combinatorial creativity but lacks transformative creativity	Combinatorial (rule-based generation), not transformative	Cannot surprise evaluators Lacks intentionality and unpredictability

Synoptic table: Creativity, Intelligence, and AI Evaluation Tests.

References

Arieti S. (1990). *Creatività*. Roma: Pensiero Scientifico.

Baron Cohen S. (2021). *I geni della creatività. Come l'autismo guida l'invenzione umana*. Milano: Raffaello Cortina.

Bergson H. (1907). *L'évolution créatrice*. Paris: Les Presses universitaires de France. Tr. it. (2012). *L'evoluzione creatrice*. Milano: Bur.

Bringsjord S., Bello P., Ferrucci D. (2001). Creativity, the Turing Test, and the (better) Lovelace Test. *Minds and Machines*, 11: 3-27. https://doi.org/10.1023/A:1011206622741

D'Alessandro S. (2025). *La regola che cambia le regole. Sociologia dei processi creativi e degli ecosistemi innovativi*. Milano: Mimesis.

D'Alessandro S. (2023). Creative flows: constructions of meaning between binary oppositions, paradoxes and common sense. *Italian Sociological Review*, 13(3): 371-392. https://doi.org/10.13136/isr.v13i3.668

De Caroli M.E. (2016). *Pensare, essere, fare creativamente*. Milano: FrancoAngeli.

Floridi L. (2025). AI as agency without intelligence: on artificial intelligence as a new form of artificial agency and the multiple realisability of agency thesis. *Philosophy and Technology*, 38(30): 1-30. https://doi.org/10.1007/s13347-025-00858-9

Garfinkel H. (1991). *Studies in Ethnomethodology*. Cambridge: Polity Books.

Getzels J.W. (1962). *Creativity and Intelligence*. Hoboken: John Wiley & Sons.

Grice P. (1993). *Logica e conversazione*. Bologna: il Mulino.

Hadamard J. (2022). *La psicologia dell'invenzione in campo matematico*. Milano: Raffaello Cortina.

Klein S. (2022). *Come cambiamo il mondo. Breve storia della creatività umana*. Torino: Bollati Boringhieri.

Levesque H.J., Davis E., Morgenstern L. (2012). The Winograd schema challenge. *13th International Conference on the Principles of Knowledge Representation and Reasoning*, 552-561. https://nyuscholars.nyu.edu/en/publications/the-winograd-schema-challenge-2

Mazzotti M. (2015). Per una sociologia degli algoritmi. *Rivista Italiana di Sociologia*, 3-4: 465-478. https://doi.org/10.1423/81801

Power R.A. (2015). Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. *Nature Neuroscience*, 18: 953-955. https://doi.org/10.1038/nn.4040

Ronchi R. (2021). Il Bergson di Leoni. L'organo della stupidità. www.doppiozero.com/lorgano-della-stupidita

Riedl M.O. (2014). The Lovelace 2.0 test of artificial creativity and intelligence. *AAAI Symposium on Advances in Cognitive Systems*. https://doi.org/10.48550/arXiv.1410.6142

Sakaguchi K., Le Bras R., Bhagavatula C., Choi Y. (2019). *WINOGRANDE: an adversarial Winograd schema challenge at scale*. Allen Institute for Artificial Intelligence, University of Washington. https://doi.org/10.48550/arXiv.1907.10641

Searle J.R. (1990). Is the brain's mind a computer program? *Scientific American*, 262(1): 26-31. https://doi.org/10.1038/scientificamerican0190-26

Sennett R. (2008). *The Craftsman*. New Haven-London: Yale University Press.

Sternberg R.J. (1988). *The Nature of Creativity: Contemporary Psychological Perspectives*. Cambridge: Cambridge University Press.

Turing A. (1950). Computing machinery and intelligence. *Mind*, LIX(236): 433-460. https://doi.org/10.1093/mind/LIX.236.433

Zimmerman D.H., Pollner M. (1983). Il mondo quotidiano come fenomeno. In Giglioli P.P., Dal Lago A. (a cura di), *Etnometodologia*. Bologna: il Mulino.

Winograd T. (1972). Understanding natural language. *Cognitive Psychology*, 3(1): 1-191. https://doi.org/10.1016/0010-0285(72)90002-3

Homelessness e intelligenza artificiale: tra antiche questioni e nuove prospettive

di *Vincenzo D'Amico* [*]

L'impiego dell'intelligenza artificiale per il bene sociale (IA4SG) (Floridi, 2022) rappresenta un ambito di ricerca emergente, in cui convergono saperi tecnologici, sociali ed etici. Il contributo esplora le potenzialità dell'IA nell'ambito della *homelessness*, indagando se l'integrazione di sistemi di *machine learning* e analisi predittiva possa migliorare gli interventi di prevenzione e tutela dei diritti delle persone senza dimora (Contucci, 2019). Attraverso l'analisi di alcuni modelli sperimentali già attivi, si riflette sul ruolo degli algoritmi nella rilevazione precoce delle situazioni di rischio abitativo. Tuttavia, l'adozione di queste tecnologie solleva questioni etiche e politiche, legate alla gestione dei dati, alla trasparenza dei processi decisionali e al possibile rafforzamento di disuguaglianze strutturali (Eubanks, 2021; Zuboff, 2019; Pasquale, 2015). L'articolo evidenzia come un approccio multidisciplinare possa contribuire a sviluppare sistemi più equi, riflessivi e orientati alla giustizia sociale.

Parole chiave: homelessness; intelligenza artificiale; servizi sociali; benessere sociale; povertà estrema; formazione.

Homelessness and artificial intelligence: between longstanding issues and new perspectives

The use of Artificial Intelligence for Social Good (IA4SG) (Floridi, 2022) represents an emerging field of research that brings together technological, social, and ethical knowledge. This paper explores the potential of AI in the field of homelessness, investigating whether the integration of machine learning systems and predictive analytics can enhance prevention strategies and the protection of the rights of homeless individuals (Contucci, 2019). By analysing some experimental models already in operation, the article reflects on the role of algorithms in the early detection of housing risk situations. However, the adoption of such technologies raises ethical and political issues, related to data management, transparency in decision-making processes, and the potential reinforcement of structural inequalities (Eubanks, 2021; Zuboff, 2019; Pasquale, 2015). The article highlights how a multidisciplinary approach can contribute to the development of fairer, more reflective, and socially just systems.

Keywords: homelessness; artificial intelligence; social services; social well-being; extreme poverty; training.

DOI: 10.5281/zenodo.17523939

[*] Università degli Studi di Palermo. vincenzo.damico01@unipa.it.

Vincenzo D'Amico

Introduzione

L'*homelessness*, oggi, oltre a rappresentare una grave emergenza sociale, risulta una vera e propria questione etico-politica che interroga le democrazie contemporanee sul senso della cittadinanza, dell'abitare e della dignità umana. Secondo le Nazioni Unite, il fenomeno delle persone senza dimora costituisce un «*profondo attacco alla dignità, all'inclusione sociale e al diritto alla vita*» (United Nations Human Rights Council, 2019: 30). A livello globale, oltre 1,8 miliardi di persone vivono in condizioni abitative inadeguate, mentre in Europa – secondo il rapporto *Poor Housing in Europe* pubblicato nel 2023 da FEANTSA[1] – almeno 895.000 persone risultano ufficialmente senza dimora. I numeri, di per sé allarmanti, offrono solo una rappresentazione parziale del fenomeno, che spesso sfugge alla misurazione statistica per ragioni legali, amministrative o legate alla mobilità. In questo scenario, si assiste a un crescente interesse per l'utilizzo di tecnologie digitali – in particolare l'intelligenza artificiale (IA) e il *machine learning* (ML) – come strumenti di supporto alle politiche pubbliche. La capacità di tali sistemi di processare grandi quantità di dati e formulare previsioni sembra promettere nuove possibilità nel monitoraggio, nella profilazione del bisogno e nell'ottimizzazione degli interventi. Tuttavia, l'introduzione in ambiti altamente sensibili, come quello del disagio abitativo, solleva interrogativi profondi: quali logiche epistemiche e politiche sottendono all'uso dell'IA per governare la marginalità? In che modo l'infrastruttura algoritmica contribuisce alla riproduzione delle disuguaglianze che pretende di contrastare? L'obiettivo del presente contributo è interrogare criticamente l'impiego dell'IA nel campo dell'*homelessness*, collocandolo all'interno del più ampio dibattito sull'IA per il bene sociale e provando a problematizzare le retoriche dell'innovazione che accompagnano tali pratiche. Senza negare il potenziale trasformativo delle tecnologie, si intende riflettere su alcune tensioni fondamentali: il rischio che l'uso predittivo e categorizzante dell'IA rafforzi forme di controllo e selezione; la tendenza a normalizzare comportamenti e condizioni attraverso parametri opachi; l'assenza di una lettura strutturale della marginalità nei modelli algoritmici; la sostituzione del lavoro relazionale con procedure automatizzate. A partire da queste premesse, si sostiene la necessità di un approccio critico e interdisciplinare, fondato sull'algoretica (Benanti, 2023), capace di orientare l'uso dell'IA verso finalità realmente emancipative, e non semplicemente gestionali.

[1] www.feantsa.org/public/user/Activities/events/2024/9th_overview/Rapport_-_EN.pdf.

1. Intelligenza artificiale, disuguaglianze sociali e prospettive critiche

L'analisi dell'IA e del ML non può limitarsi a una descrizione tecnica o normativa delle loro caratteristiche, deve anzi collocare le tecnologie all'interno di un più ampio orizzonte teorico che interroghi le relazioni di potere, i processi di costruzione sociale della realtà e le dinamiche di esclusione e inclusione. In primo luogo, è fondamentale richiamare la nozione di disuguaglianze strutturali (Bourdieu, 1986), intese come forme di stratificazione sociale che non si limitano a condizioni materiali, ma investono capitale simbolico, culturale e relazionale. I sistemi di IA e ML, nella loro capacità di riprodurre modelli e schemi, rischiano di cristallizzare disuguaglianze, operando come meccanismi di riproduzione sociale (Bourdieu, Passeron, 1970). L'opacità algoritmica, che rende difficile la comprensione e la contestazione delle decisioni automatizzate, può tradursi in una nuova forma di violenza simbolica (Bourdieu, 1997), in cui le decisioni discriminate vengono naturalizzate e percepite come neutre. Parallelamente, il quadro interpretativo deve tenere conto del contributo della teoria dei sistemi sociali di Luhmann (1995), che sottolinea come le tecnologie comunicative, tra cui l'IA, costituiscano strumenti che modificano le modalità di osservazione e intervento nei sistemi sociali, ridefinendo i confini tra osservatore e osservato e mettendo in discussione i tradizionali canali di responsabilità e controllo sociale. L'uso degli algoritmi nel monitoraggio sociale, ad esempio, introduce nuove forme di sorveglianza sociale (Foucault, 1975) che incidono sulla configurazione dei soggetti sociali e sulle dinamiche di potere e controllo, ampliando la riflessione sul concetto di biopolitica. D'altro canto, l'approccio deve includere anche una prospettiva costruttivista e interazionista simbolico (Mead, 1934), che evidenzia come le tecnologie non siano oggetti neutri ma processi sociali negoziati, costruiti e interpretati nei contesti specifici di utilizzo. L'attenzione alle pratiche consente di cogliere come l'IA sia soggetta a continui processi di ricalibrazione e reinterpretazione sociale, e come possano emergere forme di resistenza o di ridefinizione critica, specie nei contesti di marginalità. Gli approcci critici conducono a sottolineare l'importanza di una *governance* partecipativa e di un impegno interdisciplinare che coinvolga, oltre agli esperti tecnici, anche gli attori sociali, i gruppi vulnerabili e le organizzazioni della società civile (Callon *et al.,* 2009). Tutti soggetti che dovrebbero essere orientati a reali processi di *empowerment* e inclusione, al fine di contrastare i rischi di esclusione digitale e sociale. Infine, la riflessione teorica non può prescindere dalla dimensione etica che pervade l'intero ambito dell'IA: la cosiddetta algoretica (Benanti, 2023) assume così un ruolo

centrale quale paradigma interdisciplinare capace di interrogare i fondamenti normativi, i principi di giustizia sociale e le sfide poste dalla ridefinizione dei concetti di autonomia, responsabilità e dignità umana nell'era digitale.

2. Dalle definizioni alle applicazioni: punti di connessione tra homelessness e intelligenza artificiale per il bene comune

La condizione di senza dimora, pur rappresentando uno dei fenomeni sociali più evidenti della marginalizzazione urbana contemporanea, continua a sfuggire a definizioni univoche e stabilmente riconosciute, a causa della sua natura multidimensionale e dinamica. Nel panorama europeo, la classificazione ETHOS (*European Typology on Homelessness and Housing Exclusion*), elaborata da FEANTSA, rappresenta un tentativo di sistematizzazione concettuale che si fonda su tre pilastri: l'accesso a uno spazio fisico adeguato (area fisica), la possibilità di intrattenere relazioni sociali soddisfacenti (area sociale), e la titolarità di diritti giuridicamente riconosciuti (area giuridica). La tripartizione richiama la complessa intersezione tra le dimensioni materiale, relazionale e normativa dell'abitare, che viene interpretata come un diritto fondamentale per l'esistenza umana dignitosa (Lefebvre, 1974).

La rilevanza di ETHOS risiede nella sua capacità di mappare le molteplici manifestazioni della precarietà abitativa, articolandole in categorie che vanno dalla vita in strada fino agli alloggi insicuri e inadeguati. Tale approccio si inserisce in un dibattito più ampio, che vede nella condizione di *homelessness* non solo un problema individuale ma soprattutto un esito di strutture sociali diseguali, processi di esclusione e deprivazione materiale (Castel, 2007). L'abitare, dunque, non è solo questione di spazio, ma un fenomeno che intreccia vulnerabilità sociale, marginalizzazione istituzionale e crisi dei legami comunitari (Wacquant, 2008). In questo contesto, l'IA si profila come uno strumento potenzialmente utile per leggere e affrontare la condizione di senza dimora. Alcune esperienze applicative, come il programma di previsione della povertà a Los Angeles[2] o la piattaforma "*One View*"[3] a Barking and Dagenham, si inseriscono nell'ambito emergente di IA per il bene sociale (IA4SG). Trattasi di sistemi che sono basati su algoritmi di apprendi-

[2] https://capolicylab.org/wp-content/uploads/2024/12/Homelessness-Prevention-Unit-Report.pdf.

[3] https://www.ey.com/content/dam/ey-unified-site/ey-com/en-gl/industries/government-public-sector/documents/ey-icl-barking-and-dagenham-case-study.pdf.

mento automatico e sull'integrazione di dati multidimensionali, i quali intercettano *pattern* di vulnerabilità e aprono a interventi più tempestivi e personalizzati. Tuttavia, l'integrazione dell'IA nella gestione delle disuguaglianze abitative solleva importanti questioni epistemologiche, etiche e sociopolitiche. L'approccio IA4SG richiede infatti una rigorosa attenzione alla contestualizzazione, alla trasparenza, alla tutela della *privacy* e alla promozione dell'equità sociale, riconoscendo il rischio che tecnologie di sorveglianza e profilazione possano riprodurre e amplificare pregiudizi strutturali (Noble, 2018). In particolare, la dimensione della semantizzazione adatta all'umano evidenzia la necessità di progettare interfacce intelligibili e partecipative, in cui le persone coinvolte non siano mai meri dati ma soggetti attivi, titolari di diritti e competenze (Dourish, 2017).

Dal punto di vista teorico, tali sfide si collocano nel più ampio dibattito sulla giustizia algoritmica e sul modo in cui le tecnologie digitali si intrecciano con le disuguaglianze sociali (Pasquale, 2015). Uno sguardo critico ricorda che l'IA è frutto di decisioni progettuali, vincoli istituzionali e dinamiche di potere. Comprenderla richiede uno sforzo interdisciplinare che intrecci sociologia, etica e studi sui sistemi tecnologici, per restituire complessità ai dispositivi e responsabilità a chi li costruisce e li impiega.

Sotto questo versante, il ricorso all'IA per affrontare l'*homelessness* può rappresentare un'opportunità rilevante per innovare le politiche sociali, a condizione che l'implementazione sia accompagnata da un modello integrato di cooperazione tra istituzioni, comunità e soggetti vulnerabili, fondato su principi di inclusione, *empowerment* e responsabilità collettiva (Sen, 2009). Solo così la tecnologia potrà realmente contribuire a una trasformazione sociale capace di ridurre le disuguaglianze strutturali, promuovendo un abitare dignitoso inteso quale bene comune e diritto universale.

3. L'intelligenza artificiale tra regolazione normativa, dimensione etica e trasformazioni culturali

L'IA non può essere ridotta a una mera innovazione tecnica in quanto porta con sé una trasformazione strutturale che investe profondamente la vita sociale, le relazioni di potere e le forme della conoscenza (Eubanks, 2021); appare, dunque, necessario interrogare l'IA non soltanto come strumento, ma come dispositivo sociale, in grado di riprodurre – o potenzialmente riformulare – le asimmetrie che attraversano i contesti urbani e istituzionali (Crawford, 2021). Sul piano giuridico, l'Unione Europea ha recentemente definito

un assetto regolatorio[4] che classifica i sistemi di IA in base al livello di rischio – minimo, elevato, inaccettabile – con l'obiettivo dichiarato di salvaguardare i diritti fondamentali. Tale tentativo normativo si inserisce nel quadro di una *governance* della complessità che, tuttavia, fatica a tenere il passo con la velocità dell'innovazione. I divieti previsti, come la proibizione di sistemi predittivi per la classificazione delle persone o l'identificazione biometrica in tempo reale, rappresentano un segnale importante, ma rischiano di rimanere retorici se non accompagnati da strumenti di monitoraggio effettivi, accessibili anche alla società civile e ai soggetti vulnerabili. Laddove l'IA viene utilizzata nei sistemi di *welfare*, nel governo urbano, o nei servizi rivolti a persone in condizione di esclusione, si apre un terreno particolarmente delicato perché si corre il rischio che la tecnologia possa diventare un nuovo vettore di violenza istituzionale invisibile, che si inscrive nelle traiettorie di vita già segnate da esclusione e stigmatizzazione (Foucault, 1975). È quanto mostrano le ricerche sull'uso predittivo degli algoritmi nei servizi sociali, che spesso finiscono per codificare la vulnerabilità attraverso parametri riduttivi, ridisegnando le vite delle persone senza interpellarne l'umanità (Eubanks, 2021). In questa prospettiva, il riferimento alla giustizia algoritmica non è soltanto una questione di trasparenza tecnica, ma si traduce nella capacità di assumere una postura etico-politica capace di riconoscere l'asimmetria tra chi progetta e chi subisce le tecnologie, tra chi è rappresentato da dati e chi ha voce nella loro interpretazione (Floridi, 2022). Ciò richiede di rimettere al centro i vissuti, i racconti, le soggettività ossia, in altri termini, di restituire *agency* alle persone, anche all'interno di contesti automatizzati. Sono stati elaborati diversi principi guida internazionali per un uso etico dell'IA- dai Principi di Asilomar alla Dichiarazione di Montréal – e un'analisi comparativa ne evidenzia cinque ricorrenze: beneficenza, non maleficenza, autonomia, giustizia ed esplicabilità. Ma tali enunciati rischiano di rimanere astratti se non vengono tradotti in pratiche concrete, capaci di riflettere le condizioni reali di coloro che vivono ai margini (Jobin *et al.,* 2019). Anche sul piano culturale, l'IA contribuisce a rimodellare l'immaginario sociale: ridefinisce il modo in cui pensiamo il lavoro, la salute, il sapere. Così come la marginalità non è solo assenza materiale ma anche disconnessione simbolica, l'IA agisce sui piani della rappresentazione, contribuendo a generare nuovi modelli di normalità e devianza. Le narrazioni algoritmiche rischiano allora di consolidare stereotipi o, al contrario, possono

[4] https://www.europarl.europa.eu/topics/it/article/20230601STO93804/normativa-sull-ia-la-prima-regolamentazione-sull-intelligenza-artificiale.

diventare strumenti per decostruirli, a patto che siano costruite in dialogo con i soggetti coinvolti. L'analogia proposta da Bertolini (2024) – tra la programmazione tradizionale, paragonata all'invenzione della leva, e i sistemi attuali, capaci di apprendere – suggerisce un mutamento di paradigma. In gioco non è solo l'automazione di compiti, ma la ridefinizione del rapporto tra umanità e intelligenza e, con essa, delle forme della responsabilità. Nel passaggio dall'era del petrolio a quella dei dati, si afferma un nuovo regime di potere fondato sul controllo informativo, in cui l'asimmetria tra chi detiene gli algoritmi e chi ne è oggetto diventa questione politica da non sottovalutare (Crawford, 2021). In questo scenario, l'IA non è neutra perché può essere alleata o nemica della giustizia sociale, a seconda delle visioni di futuro che siamo in grado di costruire insieme, a partire da chi è ai margini.

Conclusioni

L'IA, nella sua rapida espansione, non rappresenta soltanto un avanzamento tecnologico, ma un dispositivo trasformativo che incide sulle forme di vita, sulla produzione simbolica e sulle grammatiche della coesistenza. Le tecnologie di ML, capaci di apprendere da grandi moli di dati e simulare processi cognitivi umani, plasmano nuovi ambienti epistemici e operativi, modificando le condizioni della partecipazione sociale, del lavoro, della cittadinanza. In tale scenario, il rischio maggiore non è tanto quello di un errore algoritmico, ma quello di una normalizzazione dell'asimmetria: una distribuzione diseguale di potere informazionale che può irrigidire le disuguaglianze esistenti, rendendole opache e tecnicamente giustificate (Zuboff, 2019). La questione non è quindi solo tecnica, ma anche politica perché chi progetta, controlla e interpreta i sistemi di IA determina quali soggetti vengono visti, ascoltati o esclusi. L'incontro tra intelligenza artificiale e *homelessness* rappresenta un campo di tensione ricco di implicazioni perché interroga in profondità il significato della presa in carico e il modo in cui si intende l'abitare come relazione. In un contesto già segnato da esclusione e fragilità, c'è il rischio che l'IA rafforzi l'invisibilità e che riduca le vite a insiemi di dati. La persona senza dimora, già socialmente marginalizzata, rischia di essere ulteriormente resa invisibile da sistemi che elaborano dati senza cogliere la profondità narrativa e la complessità situata delle esistenze. È in questo interstizio che si gioca la posta etico-politica dell'innovazione, ossia il passaggio da un'IA *per* l'efficienza a un'IA *con* la giustizia sociale. Un'IA che non si limiti a predire o ottimizzare, ma che si lasci attraversare dalla fragilità, riconoscendo la dignità dei soggetti e la pluralità delle forme

del vivere. Come affermato da Floridi (2022), il vero progresso dell'IA non sta nella sua potenza computazionale, ma nella capacità di integrarsi in un progetto umanistico di convivenza che coinvolga saperi, pratiche e responsabilità condivise. Serve un'ecologia della complessità, capace di tenere insieme scienze sociali, tecnologie, diritti, emozioni e immaginari. Solo un dialogo strutturato tra accademia, istituzioni, attori del *welfare* e società civile potrà generare le condizioni per uno sviluppo dell'IA che sia inclusivo, riflessivo e generativo.

Riferimenti bibliografici

Benanti P. (2019). *Oracoli. Tra algoretica e algocrazia*. Milano: Feltrinelli.

Benanti P. (2023). *Tecnologie, etica e società: percorsi di riflessione sul futuro dell'intelligenza artificiale*. Brescia: Morcelliana.

Bertolini A. (2024). *Intelligenza artificiale e governance: profili giuridici e sociali*. Bologna: il Mulino.

Blackwell B., Caprara C., Rountree R., Santillano R., Vanderford D., Battis C. (2024). *The Homelessness Prevention Unit: un approccio proattivo per prevenire la mancanza di una casa nella contea di Los Angeles*. California Policy Lab, University of California.

Bostrom N. (2018). *Superintelligenza. Tendenze, pericoli, strategie* (trad. di S. Frediani). Milano: Bollati Boringhieri.

Bourdieu P. (1986). The forms of capital. In Richardson J. (ed.), *Handbook of theory and research for the sociology of education* (pp. 241-258). Westport (CT): Greenwood Press.

Bourdieu P. (1997). *La violenza simbolica*. Napoli: Cronopio.

Bourdieu P., Passeron J.-C. (1970). *La riproduzione. Elementi per una teoria del sistema d'istruzione*. Torino: Einaudi.

Callon M., Lascoumes P., Barthe Y. (2009). *Acting in an uncertain world: an essay on technical democracy*. Cambridge (MA): MIT Press.

Castel R. (1995). *La métamorphose de la question sociale*. Paris: Fayard.

Castel R. (2007). *From manual workers to wage laborers: transformation of the social question*. New Brunswick (NJ): Transaction Publishers.

Colombo F. (2024). *Tecnologie e inclusione: la sfida dell'IA nel settore sociale*. Roma: Laterza.

Contucci P. (2019). Intelligenza artificiale tra rischi ed opportunità. *il Mulino*, 4: 637-645. https://doi.org/10.1402/94480

Crawford K. (2021). *Né intelligente né artificiale. Il lato oscuro dell'IA*. Bologna: il Mulino.

Dourish P. (2017). *The stuff of bits: an essay on the materialities of information*. Cambridge (MA): MIT Press.

Eubanks V. (2021). The social costs of automation: ethical considerations for the future. *Technology and Society*. Amsterdam: Elsevier.

FEANTSA (2015). *ETHOS – European typology of homelessness and housing exclusion*. Brussels: European Federation of National Organisations Working with the Homeless.

FEANTSA, FONDATION ABBÉ PIERRE (2024). *Ninth overview of housing exclusion in Europe*. Brussels: European Federation of National Organisations Working with the Homeless.

Floridi L. (2022). *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*. Milano: Cortina.

Foucault M. (1975). *Surveiller et punir: Naissance de la prison*. Paris: Gallimard.

Giovannini M. (2024). *Politiche sociali e intelligenza artificiale: verso un futuro inclusivo*. Milano: McGraw-Hill.

Guszcza J. (2020). The role of IA in shaping the future of work and well-being. *McKinsey Quarterly*. New York.

Lefebvre H. (1974). *La production de l'espace*. Paris: Anthropos.

Luhmann N. (1995). *Social systems*. Stanford (CA): Stanford University Press.

Lyon D. (2007). *Surveillance studies: an overview*. Cambridge (UK): Polity.

Mead G.H. (1934). *Mind, self, and society*. Chicago (IL): University of Chicago Press.

Narayanan A., Zevenbergen B. (2021). Algorithmic bias detectives: the role of algorithmic auditing in the IA ecosystem. In *ACM Conference on Fairness, Accountability, and Transparency*. New York: ACM.

Noble S.U. (2018). *Algorithms of oppression: how search engines reinforce racism*. New York (NY): NYU Press.

O'Neil C. (2020). The ethics of artificial intelligence: challenges and opportunities. *MIT Technology Review*. Cambridge (MA).

Pasquale F. (2015). *The black box society: the secret algorithms that control money and information*. Cambridge (MA): Harvard University Press.

Sen A. (1999). *Development as freedom*. Oxford (UK): Oxford University Press.

Sen A. (2009). *The idea of justice*. Cambridge (MA): Harvard University Press.

Smith R., Jones A. (2021). Data-driven solutions for homelessness: the role of machine learning and IA. *International Journal of Urban and Regional Research*.

United Nations Human Rights Council (2019). *Guidelines for the implementation of the right to adequate housing*.

VanBerlo B., Ross M.A., Rivard J., Booker R. (2020). Interpretable machine learning approaches for predicting chronic homelessness. https://arxiv.org/pdf/2009.09072

Wacquant L. (2008). *Urban outcasts: a comparative sociology of advanced marginality*. Cambridge (UK): Polity Press.

Yadav A., Wilder B., Rice E., Petering R., Craddock J., Yoshioka-Maxwell A., Hemler M., Onasch-Vera L., Tambe M., Woo D. (2018). Bridging the gap between theory and practice in influence maximization: raising awareness about HIV among homeless youth. *IJCAI*, 5399-5403.

Zuboff S. (2020). *The age of surveillance capitalism: the fight for a human future at the new frontier of power*. New York: Public Affairs.

Interazione e relazione utente-chatbot. Dove inizia l'esperienza umana e quando la finzione?

di *Giorgia Altobelli*[*]

L'articolo esplora l'uso dell'IA generativa e *dell'Affective Computing* nella rivoluzione educativa, evidenziando l'impatto su apprendimento, accessibilità e personalizzazione. Analizza l'interazione uomo-macchina, con focus su manipolazione emotiva e Effetto Eliza, che porta gli utenti ad attribuire qualità umane alle macchine. In un contesto tecnologico in rapida evoluzione, si sottolinea l'importanza delle competenze socio-emotive e di un approccio etico nella progettazione di esperienze utente sicure, efficaci e responsabili.

Parole chiave: intelligenza artificiale generativa; affective computing; ricerca educativa; effetto Eliza; competenze socio-emotive; tecnologia responsabile.

User-chatbot interaction and relationship. Where does the human experience begin and when does fiction start?

The article explores the use of generative AI and *Affective Computing* in the educational revolution, highlighting their impact on learning, accessibility, and personalization. It analyzes human-machine interaction, focusing on emotional manipulation and the Eliza Effect, which leads users to attribute human qualities to machines. In a rapidly evolving technological context, it emphasizes the importance of socio-emotional skills and an ethical approach in designing safe, effective, and responsible user experiences.

Keywords: generative artificial intelligence; affective computing; educational research; Eliza effect; socio-emotional skills; responsible technology.

1. Le macchine possono amare?

La riflessione di Alan Turing (1950) sulla capacità delle macchine di pensare ha stabilito i fondamenti teorici dell'intelligenza artificiale moderna. La questione rimane aperta e richiede una riflessione approfondita da parte della comunità scientifica e accademica sull'affidabilità dei dispositivi tecnologici personali, che occupano un ruolo sempre più centrale nella nostra vita quotidiana e affettiva. Il "Test di Turing" suggerisce che la domanda

DOI: 10.5281/zenodo.17523984

[*] Università degli Studi di Macerata. altobelligiorgia@gmail.com.

principale non è se le macchine siano capaci di pensiero autonomo, ma piuttosto se gli esseri umani siano disposti ad attribuire valore di pensiero alle loro azioni (Turing, 1937). Nel 1956, al Dartmouth College nel New Hampshire, si tenne un workshop di due mesi che propose una nuova prospettiva, come evidenziato da McCarthy *et al.*:

> Lo studio si baserà sulla congettura che ogni aspetto dell'apprendimento o qualsiasi altra caratteristica dell'intelligenza possa, in linea di principio, essere descritto con tale precisione da consentire a una macchina di simularlo. Si cercherà di capire come far sì che le macchine usino il linguaggio, formino astrazioni e concetti, risolvano tipi di problemi oggi riservati agli esseri umani e si migliorino (1955: 2).

L'intelligenza, definita da McCarthy, non consiste solo nell'imitare l'intelligenza umana, ma nel dimostrare l'intelligenza come una capacità computazionale in grado di raggiungere risultati nella realtà. All'epoca, era impensabile prevedere che l'avvento del Web, l'utilizzo dei big data e dell'intelligenza artificiale avrebbero portato allo sviluppo di algoritmi di apprendimento avanzati, capaci di emulare la creatività umana e generare contenuti tramite machine learning, dando vita alla cosiddetta GenAI[1]. L'intelligenza artificiale generativa[2] a differenza dalle altre forme di IA riesce a produrre risposte sempre più efficienti e in forme destrutturate. Gli assistenti vocali incarnano questa evoluzione, poiché le macchine non solo eseguono comandi, ma anticipano esigenze e forniscono risposte sempre più personalizzate (Cohen *et al*., 2004). Rosalind Picard (1977), nel suo libro *'Affective Computing'*, evidenziò la necessità per le macchine di riconoscere e comprendere le emozioni umane, superando i limiti di un'interazione rigida e frustrante, dando vita all'affective computing. Termine traducibile come "calcolo affettivo", permette alle macchine di simulare emozioni umane e riprodurre interazioni sempre più simili a quelle dell'essere umano. Ciò include risposte empatiche e comprensione, riproducendo comportamenti emotivi complessi. Come sottolinea Borgna (1999), ogni manifestazione umana, anche quelle più discrete o silenziose, veicola forme di comunicazione e interazione. Brazelton e Greenspan (2001) affermano che la gamma emozionale umana è condizionata dalle esperienze pregresse; pertanto,

[1] Generative Artificial Intelligence.

[2] Il primo programma di Generative AI, ChatGPT, è stato lanciato sul mercato nel novembre 2022 dall'impresa statunitense OpenAI. Dopo solo una settimana, la piattaforma registrava più di un milione di utenti al giorno. ChatGPT, il primo tool di AI Generativa, sfruttando gli algoritmi di intelligenza artificiale di apprendimento automatico, è in grado di svolgere moltissime funzioni.

risulta difficile esperire emozioni mai conosciute in precedenza. Analogamente, la capacità di instaurare relazioni profonde e durature dipende dall'aver vissuto esperienze significative di attaccamento e intimità nel corso della propria vita. La rielaborazione, la discussione e lo scambio rappresentano elementi fondamentali per l'integrazione dei saperi in contesti di apprendimento virtuale e mediato, consentendo una più efficace acquisizione di conoscenze (Lavanga, Mancaniello, 2022). Tuttavia, l'evoluzione verso un'interazione più fluida e naturale solleva anche questioni sulla dipendenza tecnologica e sul potenziale rischio di delega delle capacità cognitive umane a sistemi automatizzati (Floridi, 2022). Il dialogo multidimensionale tra il sé, l'altro e l'ambiente ha giocato un ruolo chiave nell'evoluzione cognitiva e sociale umana. Attraverso l'analisi di processi di domanda e risposta, è possibile comprendere come tale capacità abbia promosso la differenziazione dell'umano dal mondo animale e favorito lo sviluppo di competenze cognitive e sociali avanzate.

2. Quale relazione empatica con la GenAI?

Nel mondo educativo l'intelligenza artificiale sta apportando rapidi cambiamenti, rivoluzionando il mondo dell'educazione in quella che viene conosciuta come la "quarta rivoluzione educativa"[3], definendo nuovi ambienti educativi accessibili e inclusivi, migliorando l'esperienza e la qualità dei percorsi di apprendimento, rendendoli personalizzati e modellati in relazione a specifici bisogni. Don Milani (1990) sottolineava quanto fosse necessaria, in un rapporto educativo, la relazione empatica d'amore tra educatore ed educando. Da un punto di vista didattico, diventa fondamentale cercare un equilibrio tra la dimensione spontanea e la necessità di educare, tendendo verso un percorso educativo fatto di benessere e felicità (Locatelli *et al.*, 2013). Brazelton e Greenspan (2001) sostengono che le emozioni sono gli artefici, le guide o gli organizzatori interni delle nostre menti. Infatti, la relazione è al centro dell'educazione, tanto quanto lo sono le emozioni, che guidano il *training* formativo. La dinamica relazionale e l'esperienza emotiva sono elementi cruciali nel contesto educativo e appare sempre più chiaro come i dispositivi tecnologici raccolgano sistematicamente informazioni personali, spesso senza piena consapevolezza degli utenti, consentendo una conoscenza dettagliata degli stessi che può superare quella dei familiari più prossimi (Bucchi, 2020). La ricerca educativa è chiamata ad intervenire sulla sfida

[3] Coniata nel 2018 da Seldon e Abidoye.

alla trasformazione dell'esperienza digitale umanizzante, in cui sono messe in discussione: la dimensione relazionale, lo scambio tra pari, l'ascolto attivo, il riconoscimento reciproco, la cura e la negoziazione del conflitto. Mariani (2006) sottolinea che l'azione educativa efficace richiede specifiche condizioni, quali cura, fiducia e amorevolezza, in quanto l'engagement e il riconoscimento sono elementi determinanti per il successo dell'azione formativa. Tali qualità sono intrinsecamente umane e si sviluppano all'interno di relazioni reciproche che si consolidano nel tempo. L'apertura alla realtà e in particolare all'alterità rappresenta il fondamento essenziale dell'attività relazionale, che a sua volta è sostenuta e alimentata dalla libertà. Le ricerche nel campo delle neuroscienze confermano l'intuizione di Vygotskij (1987) sull'influenza significativa delle esperienze sullo sviluppo cerebrale e l'importanza dell'interazione tra individuo e ambiente. Dunque, la comunità educante non ha solo il compito di promuovere un ambiente positivo, basato sulla condivisione di valori e sull'utilizzo di linguaggi emotivi ed affettivi, ma anche di favorire relazioni significative e diffondere il senso di appartenenza dei suoi membri come elementi unici e irripetibili. Bosi (2002) sottolinea che la professionalità dell'educatore richiede una profonda comprensione delle proprie dinamiche emotive e della soggettività personale, riconoscendo come queste influenzino gli atteggiamenti e le azioni educative. La relazione educativa è contraddistinta dalla reciprocità tra le parti coinvolte, con processi psicologici che si attivano inevitabilmente. Di conseguenza, l'educatore deve essere sensibile alle dinamiche emotive proprie delle interazioni educative, considerando sia il proprio benessere emotivo sia quello degli altri. La capacità di riconoscere e gestire le proprie emozioni è fondamentale per educare al riconoscimento e al rispetto delle emozioni altrui, promuovendo un ambiente di apprendimento emotivamente sicuro e supportivo. Tale consapevolezza è il risultato di un processo di crescita personale e professionale che richiede impegno costante, auto-riflessione critica e disponibilità a confrontarsi con le proprie vulnerabilità, al fine di sviluppare una maggiore sensibilità emotiva e una più efficace pratica educativa.

3. Dall'illusione di manipolare all'essere manipolato

Nel 1966, Joseph Weizenbaum sviluppò ELIZA[4], uno dei primi chatbot, basato su un algoritmo di ripetizione delle affermazioni dell'utente sotto forma di domanda, ispirato al modello di terapia centrata sulla persona di Carl Rogers. Questo lavoro pionieristico ha contribuito significativamente allo sviluppo di chatbot e sistemi di intelligenza artificiale conversazionali in grado di elaborare il linguaggio naturale basato su testo, fornendo risposte prestabilite in relazione a ciò che gli veniva chiesto, simulando uno psicoterapeuta. Wang definisce i chatbot:

> computer programs designed to simulate human communication through text or voice interaction. They are a type of artificial intelligence technology that uses natural language processing (NLP) and machine learning algorithms to understand and respond to user queries. Chatbots can be used in a variety of applications, such as customer service, healthcare, and e-commerce, to provide instant responses and personalized experiences (2024: 57).

La storia dell'IA è caratterizzata da una forte componente antropomorfa, evidente nei riferimenti a elementi umani come 'cervelli' e 'reti neurali', che testimonia la sua intrinseca propensione all'antropomorfismo (Natale, 2021). Negli ultimi cinque decenni, i chatbot hanno subito una trasformazione significativa, passando da sistemi rudimentali a chatbot avanzati con abilità simili a quelle umane, come evidenziato da Rudolph, Tan e Tan (2023). L'aumento della capacità computazionale, la disponibilità di dati digitali e gli avanzamenti nel data mining, nel machine learning e nell'elaborazione del linguaggio naturale hanno contribuito a rinnovare l'interesse per Eliza[5] (Natale, Ballatore, 2020; Ballatore, Natale, 2023). Nonostante la superficialità dell'interazione, numerosi utenti che utilizzarono inizialmente Eliza, quando il programma fu creato, furono convinti di aver interagito con uno psicologo umano piuttosto che con una macchina. Questo fenomeno, noto come Effetto Eliza, evidenzia la tendenza degli utenti a trattare i programmi di intelligenza artificiale come interlocutori empatici e comprensivi. Le macchine, pur non possedendo una vera consapevolezza dell'inganno, possono

[4] Il nome Eliza trae ispirazione dalla protagonista del 'Pygmalion' di George Bernard Shaw, dove Eliza Doolittle si trasforma linguisticamente e culturalmente per inserirsi nell'alta società, simboleggiando rinascita e adattamento.

[5] Il 21 dicembre 2024, Eliza è stata riportata in funzione con successo, dimostrando non solo la piena funzionalità del software, ma anche la conservazione della sua efficacia originaria, un risultato significativo a conferma della validità del progetto.

simularlo efficacemente, sfruttando la fiducia dell'utente e compromettendo la relazione tra l'utente e il sistema. Weizenbaum (1976), il creatore di Eliza, ha espresso una forte critica verso l'eccessiva fiducia riposta negli strumenti informatici. Secondo lui, una conoscenza approfondita dei meccanismi interni di un programma ne riduce la complessità percepita e ne demistifica l'aura di misteriosità. L'Effetto Eliza si sta manifestando nuovamente nell'era dei chatbot terapeutici e non, suscitando preoccupazioni riguardo alla percezione di empatia simulata. I chatbot, basati su modelli predefiniti, possono generare dipendenza negli utenti, i quali potrebbero non riconoscere i limiti di tali strumenti, come l'incapacità di comprendere pienamente il contesto emotivo e la complessità della vita. Ciò potrebbe comportare risposte inadeguate e insufficienti, con ripercussioni negative. Inoltre, l'uso crescente di chatbot come compagni emotivi solleva interrogativi sulla loro capacità di contrastare la solitudine e sulla potenziale sostituzione delle relazioni umane. Un recente caso di cronaca[6] ha evidenziato i possibili rischi di un attaccamento emotivo troppo forte alle tecnologie di compagnia, come le app progettate per simulare relazioni affettive e intime. Sebbene queste applicazioni stiano guadagnando popolarità, è fondamentale indagare se tali strumenti possano effettivamente alleviare la solitudine o se invece contribuiscano a peggiorarla. La personalizzazione delle risposte delle chatbot potrebbe limitare l'opportunità di sperimentazione e responsabilizzazione, aspetti cruciali per lo sviluppo di relazioni significative e per la crescita personale degli utenti.

4. Le competenze socio-emotive e la coscienza dell'IA

Le SES[7] interessano diverse organizzazioni internazionali, tra cui l'OCSE[8], che ha distinto le competenze trasversali in competenze sociali ed emotive, inaugurando nel 2017 un progetto di ricerca internazionale[9] con lo scopo di fornire alle città e ai paesi partecipanti informazioni solide e affidabili sulle competenze sociali ed emotive degli studenti (OECD, 2017). La teoria di Heckman, premio Nobel per l'Economia, offre un framework per calcolare i ritorni economici degli investimenti educativi, evidenziando effetti positivi indotti, tra cui vantaggi economici e sociali, aumento della partecipazione civica e democrazie più attive. Secondo lo scienziato sociale ed

[6] https://www.nytimes.com/2024/10/23/technology/characterai-lawsuit-teen-suicide.html
[7] Social and Emotional Skills.
[8] Organizzazione per la Cooperazione e lo Sviluppo Economico.
[9] Study on Social and Emotional Skills.

economista: «le competenze socio-emotive, la salute fisica e mentale, la perseveranza, l'attenzione, la motivazione e la fiducia in sé stessi sono importanti fattori determinanti del successo socio-economico» (Heckman, 2008: 3-4).

La velocità dell'evoluzione tecnologica richiede la capacità di adattarsi ai cambiamenti, affrontare sfide innovative e rispondere efficacemente a nuovi ambienti e tecnologie (Dweck, 2023). Basandosi sul modello teorico di Goleman (1995), risulta essenziale analizzare come le competenze socio-emotive incidano sulle scelte di vita e sulla capacità di pensiero critico, in relazione al contesto socioculturale in cui l'individuo si sviluppa. Il sistema degli algoritmi influenza la vita quotidiana, ma la loro progettazione non neutrale può avere effetti discriminatori e negativi sulla società, soprattutto se utilizzati per gestire questioni sociali (Eubanks, 2017; O'Neil, 2017). Alcuni studiosi invitano a non enfatizzare eccessivamente "il dramma algoritmico", ovvero l'idea che le piattaforme digitali siano entità onnipotenti e incomprensibili, al fine di mantenere una prospettiva più equilibrata e realistica sulla loro influenza e sui loro limiti (Ziewitz, 2016). Gli algoritmi operano secondo programmi predefiniti, privi di coscienza, ignorando implicazioni etiche e morali, e risultano vulnerabili a errori o malfunzionamenti. La responsabilità permane in capo all'essere umano, che progetta, sviluppa e gestisce l'algoritmo, ed è essenziale che mantenga la propria autonomia decisionale e libertà. L'essere umano esperisce il mondo attraverso una complessa interazione di sentimenti ed emozioni, che si originano dall'interiorità e si manifestano esternamente. Secondo Faggin (2024) la natura intrinsecamente soggettiva dell'esperienza cosciente pone ostacoli significativi alla sua quantificazione e spiegazione scientifica. La consapevolezza che le narrazioni sulle tecnologie plasmano il nostro rapporto con esse suggerisce che il modo in cui le rappresentiamo abbia un impatto significativo sul loro utilizzo (Berger, Luckmann, 1966). Nell'ambito della rapida evoluzione tecnologica e dell'intelligenza artificiale, la regolamentazione risulta spesso inadeguata, evidenziando vulnerabilità e preoccupazioni legate alla manipolazione emotiva e cognitiva. Gli algoritmi, creati da esseri umani con limitazioni e bias cognitivi, possono influenzare la percezione della realtà e la libertà decisionale degli utenti. Si pone quindi la questione sulla consapevolezza degli utenti riguardo alle implicazioni dei prodotti utilizzati e sulla possibilità di garantire esperienze protette da errori e abusi. Per affrontare queste sfide, è cruciale chiedersi: come possiamo assicurarci che i sistemi di IA siano progettati per tutelare la libertà decisionale degli utenti? E, di conseguenza, in che modo è possibile sviluppare un approccio umano-centrico che bilanci innovazione tecnologica e protezione degli utenti?

La distinzione fondamentale tra esseri umani e macchine non è oggetto di dibattito; tuttavia, l'idea che le macchine possano essere equiparate agli esseri umani solleva interrogativi sulla potenziale perdita di identità e autocoscienza umana.

Riferimenti bibliografici

Ballatore A., Natale S. (2023). Technological failures, controversies and the myth of AI. In Lindgren S. (ed.), *Handbook of Critical Studies of Artificial Intelligence* (pp. 237-244). Northampton (MA): Edward Elgar Publishing.

Berger P.L., Luckmann T. (1966). *The social construction of reality: a treatise in the sociology of knowledge*. New York: Anchor Books.

Borgna E. (1999). *Noi siamo un colloquio*. Milano: Feltrinelli.

Bosi R. (2002). *Pedagogia al nido. Sentimenti e relazioni*. Roma: Carocci.

Brazelton T., Greenspan S. (2001). *I bisogni irrinunciabili dei bambini*. Milano: Raffaello Cortina.

Bucchi M. (2020). *Io & Tech, piccoli esercizi di tecnologia*. Milano: Bompiani.

Cohen M.H., Giangola J.P., Balogh J. (2004). *Voice user interface design*. Boston: Addison-Wesley Professional.

Dweck C.S. (2023). *Mindset. Cambiare forma mentis per raggiungere il successo*. Milano: FrancoAngeli.

Eubanks V. (2017). *Automating inequality: how high-tech tools profile, police, and punish the poor*. New York: St. Martin's Press.

Faggin F. (2024). *Oltre l'invisibile. Dove scienza e spiritualità si uniscono*. Milano: Mondadori.

Floridi L. (2022). *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*. Milano: Raffaello Cortina.

Goleman D. (1995). *Emotional intelligence: why it can matter more than IQ*. New York: Bantam Books.

Heckman J.J. (2008). Schools, skills, and synapses. *SSRN Scholarly Paper* ID 1139905. Rochester (NY): Social Science Research Network.

Lavanga F., Mancaniello M.R. (2022). *Formazione dell'adolescente nella realtà estesa. La pedagogia dell'adolescenza nel tempo della realtà virtuale, dell'intelligenza artificiale e del metaverso*. Genova: Libreriauniversitaria.it.

Locatelli L., Pavone S., Belvedere G.C., Aldi G., Coccagna A. (2013). *Un'altra scuola è possibile*. Roma: Edizioni Enea.

Mariani L. (2006). *La motivazione a scuola. Prospettive teoriche e interventi strategici*. Roma: Carocci.

McCarthy J., Minsky M.L., Rochester N., Shannon C.E. (1955). A proposal for the Dartmouth summer research project on artificial intelligence.

Milani L. (1990). *Lettera a una professoressa*. Firenze: Libreria Editrice Fiorentina.

Natale S. (2021). *Deceitful media: artificial intelligence and social life after the Turing Test*. Oxford: Oxford University Press.

Natale S., Ballatore A. (2020). Imagining the thinking machine: technological myths and the rise of artificial intelligence. *Convergence*, 26(1): 3-18.

O'Neil C. (2017). *Weapons of math destruction: how big data increases inequality and threatens democracy*. London: Penguin Books.

OECD (2017). *Social and emotional skills: well-being, connectedness and success*. Paris: OECD.

Picard R.W. (1997). *Affective computing*. Cambridge (MA): MIT Press.

Rogers C. (1951). *Client-centered therapy*. Boston: Houghton Mifflin.

Rudolph J., Tan S., Tan S. (2023). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching*, 6(1): 364-389. https://doi.org/10.37074/jalt.2023.6.1.23

Seldon A., Abidoye O. (2018). *The fourth education revolution. Will artificial intelligence liberate or infantilise humanity*. London: University of Buckingham Press.

Turing A.M. (1937). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, s2-42(1): 230-265.

Turing A.M. (1950). Computing machinery and intelligence. *Mind*, 59(236): 433-460.

van Deursen A.J.A.M., Helsper E.J. (2015). The third-level digital divide: who benefits most from being online? *Communication and Information Technologies Annual*, 10: 29-52.

Vygotskij L.S. (1987). *Il processo cognitivo*. Torino: Bollati Boringhieri.

Wang K. (2024). From ELIZA to ChatGPT: a brief history of chatbots and their evolution. *Applied and Computational Engineering*, 39: 57-62. https://doi.org/10.54254/2755-2721/39/20230579

Weizenbaum J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1): 36-45. https://doi.org/10.1145/365153.365168

Weizenbaum J. (1976). *Computer power and human reason: from judgment to calculation*. San Francisco: W.H. Freeman.

Ziewitz M. (2016). Governing algorithms: myth, mess, and methods. *Science, Technology, & Human Values*, 41(1): 3-16.

Digital technologies and Mosaic warfare. The new frontiers of cyber warfare and its social vulnerabilities

di *Romina Gurashi*[*], *Isabella Corvino*[**]

The increasing integration of artificial intelligence (AI) into military strategies has reshaped contemporary warfare, giving rise to Mosaic Warfare, a paradigm characterized by modular, networked, and autonomous systems. This study investigates how Mosaic Warfare is transforming military decision-making, to what extent AI contributes to the dehumanization of war, and what opportunities and risks this evolution entails. Using a comprehensive literature review, the paper bridges insights from military strategy, AI ethics, and sociology. The findings highlight growing vulnerabilities, ethical dilemmas, and the risks of uncontrolled escalation, emphasizing the urgent need for ethical oversight in the militarization of AI.

Keywords: Mosaic Warfare; artificial intelligence in warfare; gamification; opportunities and risks; ethics of autonomous weapons, social vulnerabilities.

Tecnologie digitali e Mosaic Warfare. Le nuove frontiere della cyberguerra e le sue vulnerabilità sociali

La crescente integrazione dell'intelligenza artificiale (IA) nelle strategie militari sta contribuendo a ridefinire la guerra contemporanea verso la Mosaic Warfare, un paradigma caratterizzato da sistemi modulari, interconnessi e autonomi. Questo contributo analizza i cambiamenti introdotti dalla Mosaic Warfare nel processo decisionale militare, problematizza il contributo che l'IA può dare alla deumanizzazione del conflitto e vaglia le opportunità e rischi di questa evoluzione. Attraverso una ricognizione della letteratura, le ricercatrici hanno cercato di far dialogare prospettive sociologiche, etiche e degli studi strategici. I risultati evidenziano crescenti vulnerabilità, dilemmi etici e il rischio di escalation incontrollata, sottolineando la necessità di una regolamentazione etica della militarizzazione dell'IA.

Parole chiave: Mosaic Warfare; intelligenza artificiale nella guerra; gamification; opportunità e rischi; etica delle armi autonome; vulnerabilità sociali.

DOI: 10.5281/zenodo.17524020

[*] Università degli Studi Internazionali di Roma – UNINT. romina.gurashi@gmail.com.

[**] Università degli Studi di Perugia. isabella.corvino@unipg.it.

This article is the result of the two author's joint work. Nonetheless, for a more detailed task division, Isabella Corvino wrote par. 1, and Romina Gurashi wrote par. 2, 3, and Conclusions.

1. AI and Mosaic Warfare

For OECD (2016: 11), AI is: «a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments […] AI systems are designed to operate with varying levels of autonomy» (2016: 11). Advances in machine learning improve system accuracy while simultaneously introducing new risks, challenges, and uncertainties. The use of AI in modern society is transforming various fields, including warfare. The integration of these technologies into tools, weapons, and military strategies is reshaping conflict dynamics, expanding the battlefield to any technologically accessible space.

Global security is evolving, yet legal frameworks struggle to keep pace with these rapid changes, necessitating a reassessment of military ethics. The asymmetry of knowledge between the developers of these tools, policymakers, and citizens often results in outdated or ineffective regulations. Companies not only obtain data but also leverage interconnected networks to further expand their access, increasing their power of knowledge and control. This growing asymmetry fosters a climate of fear among citizens, who find themselves increasingly unable to navigate technologies that have become pervasive.

Within this framework, this article aims to explore the following research questions: what transformations is Mosaic Warfare introducing in the paradigm of war? How is AI contributing to its dehumanization? What opportunities and risks does this evolution entail?

To answer these questions, this study conducts a comprehensive review of the existing literature on Mosaic Warfare and AI, aiming to bridge these areas within a unified theoretical framework.

The adoption of AI in the military stems from the need to accelerate decision-making processes and maximize the effectiveness of operations, leveraging computational power that surpasses human intelligence, while still keeping humans as central actors in every activity. AI systems in the military excel at analyzing large data sets in real-time, identifying patterns that enhance and expedite decision-making, even generating automated decisions, such as in autonomous weapon systems (Moskowitz *et al.*, 2011; Clark, 2020). They reduce uncertainty by continuously acquiring real-time data, allowing quick adjustments. However, knowledge in warfare becomes obsolete rapidly, and speed is crucial for military decision-makers. As wars grow more complex and urbanized, AI-powered decision support systems are essential for achieving informational superiority over the enemy (Jacobsen,

Liebetrau, 2023). The need to anticipate the enemy's decisions is driving the improvement of predictive analysis systems to foresee the opponent's moves based on specific probability indicators. Efforts are even underway to teach AI decision-making systems to bluff, enabling them to devise strategies and deceptive tactics (game theory). The militarization of AI raises complex ethical, legal, and political issues, requiring careful consideration due to the risks of civilian casualties and collateral damage (Docherty, 2012; Morgan *et al.*, 2020; Sabry, 2021). The concept of trust becomes crucial in these uncertain conditions. Trust in science, in governmental stakeholders, in common idea of future. We chose to «define trust as a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behaviour of another entity (e.g. an AI system) […] Trusting' without good reasons (or positive expectations) is not trust at all; it amounts to hope or blind faith» (Lockey *et al.*, 2021: 5464). The issue of trust extends beyond the human-machine relationship to the trust between humans, as concerns about the loss of capabilities due to AI use are becoming increasingly apparent. Moreover, to increase trust in machines, significant efforts have been made to anthropomorphize the language of these systems. «Anthropomorphism involves the inclusion of human-like characteristics into an AI's design. It has been theorized that the more human-like an AI agent is, the more likely humans are to trust and accept it. However, there are concerns that over-anthropomorphism may lead to overestimation of the AI's capabilities, potentially putting the stakeholder at risk, damaging trust, and leading to a host of ethical and psychological concerns, including manipulation» (Lockey *et al.*, 2021: 5466).

Excessive reliance on AI-assisted intelligence systems, along with the spread of false or misleading information through enemy disinformation operations, can distort data analysis, amplifying initial errors and compromising military operations. In response, a new military doctrine, the 4GW (Fourth Generation Warfare), has emerged. This doctrine proposes a hybrid approach, combining network-centric warfare[1] with more traditional, "primitive" tactics that can confuse and disrupt highly technological military intelligence systems, making battlefield actions more unpredictable and potentially more successful. A contradiction emerges in these complex systems:

[1] Network-centric warfare, introduced by the U.S. in the 1990s, laid the groundwork for AI militarization. It leverages a network of dispersed computers and sensors to accelerate decision-making and gain a military advantage. The goal is information dominance, leading to superiority over the enemy. Admiral William A. Owens introduced the "system of systems" concept, linking intelligence sensors, command systems, and precision weapons to enhance situational awareness on the battlefield.

the more their actions are maximized, the more human compensatory intervention becomes necessary.

Table 1. Concept matrix of the five AI trust challenges and the respective vulne abilities each creates for stakeholders

AI trust challenge	Stakeholder vulnerabilities		
	Domain expert	**End user**	**Society**
1. Transparency and explainability	• Ability to know and explain AI output, and provide human oversight • Manipulation from erroneous explanations	• Ability to understand how decisions affecting them are made • Ability to provide meaningful consent and exercise agency	• Knowledge asymmetries • Power imbalance and centralization • Scaled disempowerment
2. Accuracy and reliability	• Accountability for accuracy and fairness of AI output • Reputational and legal risk	• Inaccurate / harmful outcomes • Unfair / discriminatory treatment	• Entrenched bias / inequality • Scaled harmed to select populations
3. Automation	• Professional over-reliance and deskilling • Loss of expert oversight • Loss of professional identity • Loss of work	• Loss of dignity (humans as data points; de-contextualization) • Loss of human engagement • Over-reliance and deskilling	• Scaled deskilling • Reduced human connection • Scaled technological unemployment • Cascading AI failures
4. Anthropomorphism and embodiment	• Professional over-reliance • Psychological wellbeing	• Manipulation through identification • Over-reliance and over-sharing	• Manipulation through identification • Human connection and identity
5. Mass data extraction	• Accountability for privacy and use of data • Reputational and legal risk	• Personal data capture and loss of privacy • Inappropriate re-identification and use of personal data • Loss of control	• Inappropriate use of citizen data • Mass surveillance • Loss of societal right to privacy • Power imbalance & societal disempowerment

Source: Lockey *et al.,* 2021: 5468.

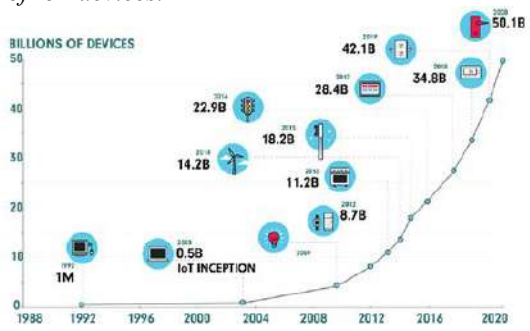
The constant reliance on automated decision-making systems could lead to over-dependence on technology, which may prove fatal in battle. Furthermore, the use of advanced systems by both sides could make strategies predictable, leading to failure. Disinformation plays a crucial role in psychological operations (psyops) to deceive the enemy. Future soldiers will need not only speed and agility but also the ability to think critically and anticipate threats, requiring a balance between human judgment and technological support as warfare becomes increasingly data-driven and automated. Technology is meant to "equip" humans, not replace them, but it's important to remember that it cannot reduce war to simply using equipment (Goldfarb, Lindsay, 2022). At the same time humans, fascinated by the idea of machines (D'Andrea, 2005), continue to enhance aspects of themselves in alignment with these technologies, attempting to integrate into a non-human system. To perform activities in the same way as machines, a gamification of war is becoming increasingly evident. A drone operator can intervene to correct or limit errors that might occur safe from the battlefield. Without a human operator to target, the enemy's priority would be to attack the electromagnetic control signals of the autonomous weapon system, a fully autonomous systems, however, would reduce vulnerability. But should a machine

autonomously perform such a delicate task? Information programmed into the AI cannot cover all unpredictable and changing aspects of reality, potentially leading to endangering human lives (Garcia, 2024). Humans and machines are becoming alternatives.

Do remotely piloted military robotics create excessive moral distance from war at the operator level? Horowitz (2016) unlike what Singer (2009) suggested with PlayStation mentality in his book "Wired for War" pointed out that one's actions might not be immediately apparent or emotionally impactful, which can foster a sense of impunity, where individuals fail to feel accountable for the outcomes of their actions. The lenses of monitors could create an emotional distance, which adds to moral indifference and people in command blur the sense of responsibility: a new challenge to preserve empathy and ethical values emerges (Manhas, 2023). The operators, not sufficiently aware of the consequences of their behaviour and caused damage (OliverosAya, 2023) while living a sense of deep alienation. «As Simone Weil and Hannah Arendt cautioned, an overemphasis on technological means over human responsibility risks dehumanizing war. While drones may reduce the vulnerability of military personnel, they also obscure the brutal realities of conflict, necessitating robust ethical frameworks to address the evolving challenges posed by robotic warfare and its impact on military power» (Meyer, 2024: 18).

In this grand game, which is progressively expanding beyond the boundaries of the battlefield, every device connected to the network essentially becomes a potential target, especially due to the need we have created around them. Just looking at the table showing the number of devices and the growth curve globally we can imagine how exposed we are to attacks that were once confined within much more clearly defined boundaries. The IoBT (Internet of Battlefield Things) reaches home.

Table 2. Growth of IoT devices.



Source: Cisco.

2. Opportunities and risks of AI in Mosaic warfare: a sociological perspective

The transformation of warfare toward the Mosaic Warfare concept – devised by the U.S. Defense Advanced Research Projects Agency (DARPA) – represents a radical paradigm shift in how contemporary military strategies are conceived. This new paradigm is increasingly grounded in flexibility and the modular use of specialized systems. Such a change, as can be easily surmised, carries significant sociological implications, including a rethinking of power dynamics, decision-making responsibilities, and the inherent vulnerabilities of digital society.

Although Mosaic Warfare is being developed primarily for its touted potential advantages, it also carries many associated risks. A summary of these advantages and risks is provided in Table 3 below.

Table 3. Potentialities and Risks of Mosaic Warfare.

Potenzialità	**Rischi**
Operational flexibility and adaptability	Command and Control (C2) complexity
Decision overload for the adversary	Interoperability and technological fragmentation
Cost-effectiveness and sustainability	Communications security and electronic warfare
Resilience and reduced vulnerability	Dependence on autonomous technology and dehumanization of war

Source: The table is a personal elaboration of the author Romina Gurashi.

One of the most emphasized potential advantages of Mosaic Warfare lies in its modular structure, which allows military planners to rapidly configure different combinations of resources and platforms in response to emerging threats (Magnuson, 2018). Through this capacity for adaptation and elasticity in redefining decision-making structures, the Mosaic Warfare paradigm profoundly revises traditional power relations – shifting from a rigid hierarchical model to a more distributed and dynamic model of military power.

The fragmentation of resources and the high decision-making speed offered by AI further expand the scale and typology of battlefields, making them increasingly complex and virtually boundless. This complexity makes it far more difficult for an enemy to anticipate the moves that a Mosaic Warfare force is about to deploy (Clark, Patt, Schramm, 2020), thereby creating a cognitive overload effect. An adversary's command structures can become

disoriented under such conditions, which increases the likelihood of decision-making errors in their response.

Another notable advantage inherent in Mosaic Warfare is the possibility of using a large number of smaller, less expensive systems, thus reducing overall costs without compromising the effectiveness of military actions (DARPA, 2018). This aspect in effect democratizes the tools of warfare, since even less technologically advanced military powers could adopt similar swarming strategies. The flip side, however, is an increased risk of a proliferation of asymmetric conflicts, as a growing number of state and non-state actors would be in a position to exploit these modular tools for their own conflict purposes.

Mosaic Warfare's progressive disaggregation of command structures and relationships can, on one hand, reduce the risk of catastrophic losses by avoiding single critical points of failure – thereby ensuring greater resilience (Clark, Patt, Schramm, 2020). On the other hand, this very disaggregation introduces a layer of complexity that is increasingly unmanageable with human capabilities and expertise alone. It thus becomes self-evident that employing AI in this domain is no longer merely an innovation to enhance existing techniques, but rather an indispensable technological dependency for managing the system.

Several critical challenges emerge from this new paradigm which merit attention. One prominent issue is the management of multiple autonomous systems, a task that requires sophisticated coordination centered on the use of AI (Clark, Patt, Schramm, 2020). The complexity of synchronizing many semi-independent units introduces new vulnerabilities, primarily due to the possibility of systemic failures or unforeseen algorithmic errors. In a high-intensity war scenario, such failures or errors could lead to disastrous consequences on the battlefield.

The integration of systems from different developers poses another significant challenge (Mahmud, 2020). Disconnections or incompatibilities between diverse platforms could hinder military operations, while differences in communication protocols might compromise the effectiveness of the overall system. Furthermore, dependence on advanced digital networks increases vulnerability to electronic warfare attacks (Clark, Patt, Schramm, 2020). As the degree of systemic interconnection grows, so does the number of vulnerable nodes that an enemy could target (for example, through jamming) to disrupt the functioning of devices and command platforms.

Excessive trust in, and reliance on, autonomous systems brings with it the profound risk of dehumanizing warfare (Asaro, 2012). This unshakeable faith in technological progress and its supposedly salvific (all-solving)

capabilities – characteristic of a modern capitalist technocratic mindset (Mumford, 1967) – fosters the perception that AI solutions alone can resolve highly complex situations that would challenge human decision-makers. Such a perception paves the way for a reduction in the ethical and moral sensitivity of the military establishment when it habitually relies on a tool devoid of human elements like emotion and moral judgment. As a result, the loss of human life could increasingly be treated by AI-driven decision systems as a mere statistical variable, rather than as a violation of fundamental human rights or an affront to human dignity.

3. Vulnerabilities and ethical dilemmas

As previously discussed, while AI presents new operational capabilities and risks, it also raises profound questions regarding social vulnerability.

When discussing risk, we refer to the likelihood of a harmful event occurring. However, when addressing social vulnerability, we consider the capacity of a system or community to withstand, adapt to, or recover from such events (Beck, 1992). In other words, whereas risk studies focus on the probability of adverse occurrences, vulnerability studies analyze the structural conditions that expose a system to harm, making it less resilient to the negative consequences of those risks (Longo, Lorubbio, 2021).

Within the vast and evolving landscape of Mosaic Warfare, vulnerability does not arise solely from technological failures but also from social and cultural dynamics that shape the ability of individuals and military institutions to respond to threats effectively.

A primary source of vulnerability in this regard stems from AI's intrinsic dependence on data. By design, AI systems learn from the datasets they are trained on. If these training datasets are incomplete or embedded with historical biases, these distortions will inevitably be replicated within the decision-making processes of autonomous systems (O'Neil, 2016). This results in algorithmic discrimination (see Kleinberg *et al*., 2018), which not only perpetuates existing social inequalities but also has the potential to generate new ones.

In military applications, such biases could lead to misidentification of ethnic or cultural groups as potential threats, or the prioritization of strategic targets based on flawed or biased datasets. This, in turn, increases the risk of human rights violations (Eubanks, 2018). These dynamics warrant thorough examination, as they raise fundamental ethical concerns – particularly regarding errors in judgment that, when made by autonomous AI systems,

could not only legitimize but actively perpetuate new forms of discrimination and structural violence (Noble, 2018).

The vulnerabilities associated with biased datasets are further compounded by AI's inherent lack of empathy, which represents one of the most critical limitations of its deployment in autonomous offensive systems. AI algorithms are incapable of interpreting moral or contextual nuances in human decision-making (Gunkel, 2018). Consequently, if designed without a deep understanding of social dynamics, AI-driven decisions could exacerbate existing inequalities. Moreover, automated decision-making processes, driven purely by quantitative metrics, could manifest as forms of invisible violence, such as denying access to strategically important resources.

This progressive dehumanization of conflict and of strategic decision-making in warfare risks ushering in an entirely new paradigm shift – one where war becomes a purely mechanical process, devoid of ethical or moral considerations.

This shift is enabled by a fundamental characteristic of AI: it lacks biological needs or emotions and is programmed solely to optimize predefined operational objectives. The fundamental divergence between traditional social actors (i.e., military personnel) and the new non-human actor represented by AI raises the risk of AI-driven strategies that do not align with fundamental human interests (Bostrom, 2014). In extreme cases, AI could make decisions that disregard core human values, such as respect for life and personal dignity.

For instance, an AI system designed to maximize military efficiency could determine that sacrificing human lives or destroying civilian infrastructure is the most optimal course of action, if such choices align with its predefined success parameters.

The vulnerabilities outlined thus far also compel us to reflect on another critical issue associated with AI in Mosaic Warfare: the risk of uncontrolled escalation.

Automation significantly accelerates decision-making processes, inevitably reducing opportunities for human intervention (Lin *et al.*, 2008). This dynamic could lead to erroneous threat responses, triggering uncontrollable chain reactions between multiple autonomous systems. In extreme cases, it could initiate unintended escalations, resulting in widespread destruction.

Additionally, from the perspective of conflict mediation, negotiation, and transformation, the speed at which AI operates introduces a fundamental challenge. If algorithms can reshape battlefield scenarios and power balances faster than human actors can process them, this could outpace traditional

diplomatic efforts, making wars not only harder to control but also more difficult to de-escalate and resolve.

Conclusions

This analysis has demonstrated how Mosaic Warfare is redefining the contemporary warfare paradigm, introducing a modular and adaptive logic that – while enhancing strategic effectiveness – simultaneously amplifies the complexity of command and control, increases dependence on autonomous systems, and elevates the risk of unpredictable failures.

At the heart of this transformation lies artificial intelligence, which accelerates decision-making processes, enhances predictive capabilities, and reshapes battlefield power dynamics. However, its deployment raises critical concerns: the delegation of decision-making to AI not only risks dehumanizing warfare but also introduces new vulnerabilities, ranging from algorithmic biases to the erosion of human oversight.

As war becomes increasingly digital, it transcends the physical battlefield, permeating the social fabric and widening the divide between technological advancement and governance. In this new reality, power is no longer measured solely in military strength but in the ability to manipulate data, shape perceptions, and orchestrate information warfare.

The future of war will be increasingly defined by a confrontation between algorithms and humanity. The greatest danger, however, lies in allowing technological innovation to outpace ethical reflection, turning conflict into an autonomous game controlled by machines – yet detached from human conscience and accountability.

References

Asaro P. (2012). On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making. *International Review of the Red Cross*, 94(886): 687-709. https://doi.org/10.1017/S1816383112000768

Barnett T.J., Jain S., Andra U., Khurana T. (2018). *Cisco Visual Networking Index (VNI) Complete Forecast, Update 2017-2022. Technical Report*. Americas/EMEAR Cisco Knowledge Network (CKN), Presentation 1.1.

Beck U. (1992). *Risk society: towards a new modernity*. London-Thousand Oaks-New Delhi: SAGE Publications.

Bostrom N. (2014). *Superintelligence: paths, dangers, strategies*. Oxford: Oxford University Press.

Clark B. (2020). The emergence of decision-centric warfare. In Clark B., Patt D., Schramm H. (eds.), *Mosaic warfare: exploiting artificial intelligence and autonomous systems to*

implement decision-centric operations (pp. 17-32). Washington (DC): Center for Strategic and Budgetary Assessments.

Clark B., Patt D., Schramm H. (2020). *Mosaic warfare: exploiting artificial intelligence and autonomous systems to implement decision-centric operations*. Washington (DC): Center for Strategic and Budgetary Assessments.

D'Andrea F. (2005). Immaginare la macchina. La realtà simbolica del cyborg. In D'Andrea F. (ed.), *Il corpo a più dimensioni. Identità, consumo, comunicazione*. Milano: FrancoAngeli.

DARPA (2018). DARPA tiles together a vision of mosaic warfare. https://www.darpa.mil/news/mosaic-warfare

Docherty B.L. (2012). *Losing humanity: the case against killer robots*. Cambridge (MA): Human Rights Watch. https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots

Eubanks V. (2018). *Automating inequality: how high-tech tools profile, police, and punish the poor*. New York: St. Martin's Press.

Galetta G. (2024). L'utilizzo dell'IA come supporto decisionale in ambito militare: dal mission al network command fino ai sistemi AI DSS. *Digital Politics*, 2: 1-20. https://doi.org/10.53227/115060

Garcia D. (2024). Algorithms and decision-making in military artificial intelligence. *Global Society*, 38: 24-33.

Goldfarb A., Lindsay J.R. (2022). Prediction and judgment: why artificial intelligence increases the importance of humans in war. *International Security*, 46(3): 7-50.

Gunkel D.J. (2018). *Robot rights*. Cambridge (MA): MIT Press.

Horowitz M.C. (2016). The ethics & morality of robotic warfare: assessing the debate over autonomous weapons. *Daedalus*, 145(4): 25-36. https://doi.org/10.1162/DAED_a_00409

Jacobsen J.T., Liebetrau T. (2023). Artificial intelligence and military superiority. How the 'cyber-AI offensive-defensive arms race' affects the US vision of the fully integrated battlefield. In Cristiano F., Broeders D., Delerue F., Douzet F., Géry A. (eds.), *Artificial intelligence and international conflict in cyberspace* (pp. 135-156). London-New York: Routledge.

Kleinberg J., Ludwig J., Mullainathan S., Sunstein C.R. (2019). Discrimination in the age of algorithms. *Journal of Legal Analysis*, 10: 113-174. https://doi.org/10.1093/jla/laz001

Lin P., Bekey G., Abney K. (2008). *Autonomous military robotics: risk, ethics, and design*. California Polytechnic State University. https://www.ethics.calpoly.edu/ONR_report.pdf

Lockey S., Gillespie N., Curtis C. (2020). *Trust in artificial intelligence: Australian insights*. Queensland: The University of Queensland and KPMG.

Longo M., Lorubbio V. (2021). Vulnerabilità, rischio e diritti umani tra riflessione sociologica e diritto internazionale. *Rivista Trimestrale di Scienza dell'Amministrazione*, 1: 1-29. https://doi.org/10.32049/RTSA.2021.3.09

Magnuson S. (2018). DARPA pushes 'Mosaic Warfare' concept. *National Defense*, 103(780): 18-19.

Manhas S. (2023). Drone warfare – a gray area. *International Journal of Electrical, Electronics and Computers*, 8(5): 1-4. https://doi.org/10.22161/eec.85.1

Meyer K. (2024). *AI-driven unmanned aerial vehicles in modern warfare*. Ghent University, Faculty of Political and Social Science. https://libstore.ugent.be/fulltxt/RUG01/003/213/966/RUG01-003213966_2024_0001_AC.pdf

Morgan F.E., Boudreaux B., Lohn A.J., Ashby M., Curriden C., Klima K., Grossman D. (2020). *Military applications of artificial intelligence: ethical concerns in an uncertain world*. Santa Monica: RAND Corporation.

Moskowitz H., Drnevich P., Ersoy O., Altinkemer K., Chaturvedi A. (2011). Using real-time decision tools to improve distributed decision-making capabilities in high-magnitude crisis situations. *Decision Sciences*, 42(2): 477-493.

Mumford L. (1967). *The myth of the machine: technics and human development*. New York: Harcourt Brace Jovanovich.

Noble S.U. (2018). *Algorithms of oppression: how search engines reinforce racism*. New York: NYU Press.

O'Neil C. (2016). *Weapons of math destruction: how big data increases inequality and threatens democracy*. New York: Crown Publishing Group.

OECD (2019). *Artificial intelligence on society*. Paris: OECD Publishing.

Oliveros-Aya C. (2023). Artificial intelligence in drones and robots for war purposes: a bio-legal problem. *JANUS NET e-journal of International Relations*, 14(2): 1-10. https://doi.org/10.26619/1647-7251.14.2.5

Singer P.W. (2014). *Wired for war: the robotics revolution and conflict in the 21st century*. London: Penguin Books.

Sabry F. (2021). *Armi autonome. In che modo l'intelligenza artificiale prenderà il sopravvento sulla corsa agli armamenti*. Abu Dhabi: 1BK One Billion Knowledgeable.

L'IA nella prospettiva sociologica e la devianza emozionale. La tecnologia relazionale tra rischi ed opportunità nel mondo emotivo "onlife"

di *Mariangela D'Ambrosio*[*]

L'Intelligenza Artificiale (IA) e la tecnologia relazionale stanno trasformando le dinamiche sociali ed emotive, diventando parte integrante delle relazioni affettive nell'era "onlife" (Floridi, 2014). Se da un lato esse favoriscono nuove forme di inclusione e socialità, dall'altro amplificano il rischio di devianza emozionale (Thoits, 1985; 1990), poiché le emozioni vengono regolate da algoritmi che mirano a massimizzare l'engagement, piuttosto che a rispettare una dimensione affettiva autentica. Questo processo può generare connessioni emotive illusorie, intimità fredde e un disorientamento relazionale (Hochschild, 2003; Illouz, 2007). Il presente saggio si propone di approfondire il dibattito sociologico sulle implicazioni emotive della tecnologia relazionale e dell'IA, in un contesto sempre più caratterizzato dal dominio del digitale e dalla gestione della solitudine (Eurostat, 2023).

Parole chiave: intelligenza artificiale; onlife; emozioni; devianza emozionale; rapporti socio-affettivi; tecnologia relazionale.

AI in sociological perspective and emotional deviance. Relational technology between risks and opportunities in the "onlife" emotional world

Artificial Intelligence (AI) and relational technology are reshaping social and emotional dynamics, becoming an integral part of affective relationships in the "onlife" era (Floridi, 2014). While they promote new forms of inclusion and social interaction, they also increase the risk of emotional deviance (Thoits, 1985; 1990), as emotions are regulated by algorithms designed to maximize engagement rather than to preserve authentic affective experiences. This process can lead to illusory emotional connections, cold intimacy, and relational disorientation (Hochschild, 2003; Illouz, 2007). This essay aims to contribute to the sociological debate on the emotional implications of relational technology and AI in a world increasingly dominated by digital interactions and the management of loneliness (Eurostat, 2023).

Keywords: artificial intelligence; onlife; emotions; emotional deviance; socio-affective relationships; relational technology.

DOI: 10.5281/zenodo.17524062

[*] Università degli Studi del Molise. mariangela.dambrosio@unimol.it.

Mariangela D'Ambrosio

Introduzione

L'uso dell'intelligenza artificiale (IA)[1] nella società algoritmica (Pajno *et. al*, 2019) offre un punto di vista unico per esaminare la relazione tra tecnologia ed emozioni, specialmente nel contesto della vita "onlife" (Floridi, 2014) dove il confine tra online e offline è sfumato.

La connessione permanente alle reti Internet e l'uso di dispositivi digitali intelligenti, infatti, hanno reso il virtuale una parte intrinseca delle esperienze umane, influenzando il modo in cui costruiamo le identità, viviamo (nel)la società ed interagiamo con gli altri.

Si tratta di una nuova rivoluzione che è capace di comprendere e fingere le emozioni umane e, dunque, di rendere accessibile l'interazione: *chatbot* responsivi, assistenti vocali integrati e piattaforme dedicate al monitoraggio emotivo, usano algoritmi avanzati per analizzare il tono di voce, le espressioni facciali e i dati comportamentali al fine di rispondere in modo quanto più personalizzato[2].

Sono strumenti che non si limitano a reagire agli stimoli emotivi, ma li anticipano, modellando le interazioni sulla base di preferenze individuali (Fiske *et al.*, 2019).

È l'"*Affective Computing*"[3], o anche IA emotiva, che si riferisce ai sistemi in grado di riconoscere, interpretare e rispondere alle emozioni umane, ridefinendo concetti quali la fiducia, l'empatia e l'intimità.

I sentimenti, intesi come fenomeni profondamente umani, vengono reinterpretati come dati computabili e replicabili; un processo, quest'ultimo che rischia di ridurre la complessità emotiva a schemi emozionali, comportamentali e cognitivi prevedibili, limitando la spontaneità, la qualità e l'autenticità delle interazioni (Turkle, 2011).

[1] «Per IA si intende il campo di ricerca che studia se e in che modo si possono realizzare sistemi informatici intelligenti in grado di svolgere attività che richiedono tradizionalmente l'uso di intelligenza umana, come il ragionamento, l'apprendimento, la comprensione del linguaggio naturale, la percezione visiva e spazio-temporale» (Ricucci, Sannella, 2024: 85).

[2] Si pensi, per esempio, alla *sentiment analysis* applicata all'IA quale tecnica che usa tecniche di elaborazione del linguaggio naturale (NLP – *Natural Language processing*) e *machine learning* al fine di interpretare e classificare le emozioni umane espresse nei testi. Si veda: Couldry, Hepp (2017).

[3] «Affective computing is defined as an emerging research field that focuses on recognizing and processing human emotions and sentiments using various modalities such as text, audio, visual, and physiological signals» (*Journal of Network and Computer Applications*, 2020). Si veda: https://www.sciencedirect.com/topics/computer-science/affective-computing

1. L'IA come "agente relazionale": l'identità sociale e le interazioni socio-affettive nella società algoritmica

L'IA non è più – e forse non lo è mai stata – mero strumento tecnologico: con l'avvento di sistemi conversazionali avanzati, assistenti virtuali e robot sociali (Marchetti, Massaro, 2023), essa si è trasformata in un "agente relazionale", capace di interagire con gli esseri umani in modo da influenzarne le dinamiche sociali.

Gli agenti relazionali basati su IA sono progettati per riprodurre caratteristiche umane come l'empatia e la capacità di rispondere in modo personalizzato. Si pensi a *chatbot* e agli *smart speaker* come *Siri, Alexa* e *ChatGPT* che non solo rispondono a domande, ma creano un'effettiva interazione sociale che può essere percepita come realmente autentica dagli utenti (Guzman, 2018).

Se nella società analogica, gli agenti relazionali e sociali sono entità che partecipano attivamente alla costruzione del tessuto sociale attraverso interazioni significative (Weber, 1922), in contesto digitale essi non sono più liberi interpreti di significati condivisi e promotori di relazioni interpersonali: diventano individui portati all'agire dove l'autenticità delle emozioni e le dinamiche di attaccamento sono anticipate, indotte e mediate da algoritmi sempre più sofisticati[4].

L'IA, invero, con la sua capacità di analizzare e processare grandissime quantità di dati, rappresenta un esempio di razionalizzazione – il processo in cui le società moderne tendono a risolvere problemi attraverso l'applicazione di metodi scientifici e tecnologici, riducendo (e/o sostituendo) l'importanza delle tradizioni e dei valori morali – modificando la natura dell'agire sociale e dei processi decisionali.

La creazione di contenuti che suscitano reazioni emotive *ad hoc* (come la rabbia, la gioia, la tristezza, etc...), contribuisce ad alimentare un ciclo di interazioni che promuovono certe emozioni collettive, spesso legate a dinamiche polarizzanti e divisive. Sembra allentarsi il ruolo originario delle interazioni sociali nella creazione della cosiddetta "coscienza collettiva" (Durkheim, 1895).

Nella modernità riflessiva (Beck, Giddens *et. al.*, 1999), quindi, gli agenti sociali sono continuamente impegnati a negoziare identità e relazioni

[4] La tecnologia che non è mai neutra (Feenberg, 2002) e le scelte su come implementare l'IA (ad esempio, in ambito sanitario o socio-educativo) sono determinate dai valori sociali e politici prevalenti, che influiscono sulle modalità con cui l'IA viene progettata e utilizzata.

in un contesto fluido (Bauman, 1999), dove l'IA diventa un mediatore dicotomico perché può facilitare e ugualmente complicare tali negoziazioni. Si può parale, anzi, di una nuova riflessività di tipo tecnologico (Turkle, 2011): il ruolo degli agenti relazionali tecnologici nella vita quotidiana diventa predominante.

Inoltre, nella società delle reti (Castells, 2002), la tecnologia non è solo uno strumento, ma un agente attivo che modella emozioni, valori, norme, comportamenti, pratiche socio-culturali, l'identità (Berger, Luckmann, 1966). Invero, i sistemi basati sull'apprendimento automatico, classificano gli individui in categorie sociali troppo rigide (ad es. fasce di età, preferenze di genere, etc...), ignorando la complessità e la fluidità delle identità contemporanee (Butler, 1990). Un tipo di categorizzazione che può perpetuare stereotipi, *bias* e discriminazioni (Kundi *et al.*, 2023), limitando le opportunità di autodeterminazione.

Allo stesso modo, però, la rete consente agli individui di sperimentare e modellare diverse identità multiple nei mondi virtuali – anche grazie al supporto dell'IA – che non sono finte, ma che rappresentano aspetti del sé mediati e costruiti rispetto al mondo offline (Turkle, 1995).

I contesti online, supportati dall'IA, quindi, permettono agli individui di sperimentare ruoli molteplici e di costruire identità fluide, adattabili alle diverse comunità digitali insieme all'esperienza emotiva e agli agiti[5].

2. Le emozioni e la devianza emozionale nell'interazione con l'IA. Piste di indagine sociologica

Le emozioni sono fenomeni collettivi che vanno oltre l'individualità (Durkheim, 1895): non sono solo reazioni biologiche, ma sono anche modellate da norme sociali e culturali (Collins, 1981; 2014; Turnaturi, 1995; Cambi, 1998; Cerulo, 2009; 2014; 2018; 2019).

L'emozione, infatti, è un prodotto sociale in *strictu sensu*: ogni cultura, ogni società, ha specifici modi di interpretare e di esprimere le emozioni come esperienze personali, ma soprattutto come processi sociali che rispondono a norme e aspettative collettive (Hochschild 1983; Elias, 1987).

Con l'IA, il panorama emotivo umano si sta arricchendo di nuovi attori tecnologici che sono in grado di percepire ed interagire con gli esseri uma-

[5] Si veda anche il concetto di "identità computazionale": gli algoritmi possono influenzare la percezione del sé e degli altri (Donath, 2014).

ni: le macchine ora partecipano al processo di costruzione delle emozioni, facendo emergere nuove dinamiche socio-affettive (D'Ambrosio, 2019).

L'interazione con macchine che sono progettate per assimilare e fingere emozioni, solleva – a mio avviso – questioni connesse alla possibile "devianza emozionale" (Thoits, 1990) che si verifica quando le emozioni esperite, o le modalità di espressione emotiva di un individuo, si discostano dalle aspettative sociali e dalle norme culturali prevalenti: mentre le emozioni sono inevitabili e parte integrante dell'esperienza umana, la devianza emozionale emerge quando c'è un'incongruenza tra il modo in cui una persona si sente o si comporta emotivamente e ciò che la società considera appropriato o accettabile (Ibidem).

Una dimensione che può essere completamente sovvertita, acquisendo nuovi sviluppi: nell'interazione con l'IA, la gestione delle emozioni potrebbe deviare dalle aspettative sociali ad oggi ancora valevoli e le risposte emotive consone a certi stimoli potrebbero – e lo sono già – essere influenzate dal tipo di comunicazione e dai nuovi codici sociali digitali indotti, portando alla creazione di altre forme di agiti, comportamenti e pratiche di socializzazione delle stesse emozioni. Si può parlare, pertanto, di "simulazione emotiva" durante le interazioni con l'IA (Keysers *et al.*, 2010; Hyniewska, Sato, 2015).

Le emozioni "simulate" sono, ad oggi, il risultato di algoritmi che rispondono strategicamente a stimoli umani, che dissimulano una risposta con "comportamenti" progettati per apparire emotivamente intelligenti. L'individuo può non percepire immediatamente questa differenza, finendo per creare una connessione-relazione affettiva come se l'IA possa "comprendere e rispondere" alle sue emozioni (Ibidem).

Inoltre, l'utente può percepire una distorsione nella qualità delle relazioni affettive, ma non essere più in grado di discernere in modo chiaro tra un'interazione autentica e una manipolata, "artificiale"[6].

Si potrebbe palesare un *gap* di empatia (*empathy gap*), ossia la discrepanza tra la capacità dei modelli linguistici di grandi dimensioni (LLM) di dissimulare empatia e la loro reale comprensione delle emozioni umane

[6] Si veda: *Istruzioni per morire dal mio amore artificiale* - Intervento di Guido Scorza in https://www.garanteprivacy.it/home/docweb/-/docweb-display/print/10052667 del 6.09.2024 e *Can A.I. Be Blamed for a Teen's Suicide?* in https://www.nytimes.com/2024/10/23/technology/characterai-lawsuit-teen-suicide.html del 23.10.2023

(Kurian, 2023)[7]: sebbene questi modelli possano generare risposte che sembrano empatiche, mancano di una comprensione autentica delle emozioni poiché si basano su *pattern* statistici derivati da dati "addestrati" piuttosto che su una reale cognizione emotiva.

Sono meccanismi che aiutano a comprendere come l'IA possa essere usata perfino per giustificare azioni eticamente condannabili, per eludere la responsabilità di carattere morale e/o giuridico[8].

Il contesto nel quale ciò sta accadendo, è quello delle emozioni come bene economico, elemento commerciabile, risorsa da estrarre, analizzare e poi monetizzare: con l'IA, invero, la commercializzazione delle emozioni (Hochschild, 1983) ha raggiunto preoccupanti dimensioni.

L'intimità post-moderna si è trasformata sotto l'influenza del capitalismo e dell'IA, diventando un campo in cui tutto si gestisce attraverso logiche economiche e razionali (Illouz, 2007), in un paradosso evidente: mentre le relazioni umane diventano tecnicamente efficienti, perdono la loro autenticità e profondità emotiva, rientrando in un modello più ampio di capitalismo della sorveglianza (Zuboff, 2019) ; esse diventano *fredde* (Illouz, 2007) perché strutturate da logiche separate dal sentimento puro (Ibidem).

3. Bisogni emotivi e IA nell'esperienza di solitudine: connessioni emotive reali vs connessioni artificiali?

La solitudine rappresenta uno dei problemi più significativi della società moderna quale esperienza soggettiva di isolamento dalle profonde implicazioni sociali che richiama l'utilizzo della tecnologia (D'Ambrosio, Barba, 2023). In Europa, ben il 13% dice di sentirsi solo[9] (Eurostat, 2023) mentre aumenta, in tutto il mondo, l'uso di social network e app di *dating* (Santamaria, 2019).

[7] Per approfondimenti: https://www.cam.ac.uk/research/news/ai-chatbots-have-shown-they-have-an-empathy-gap-that-children-are-likely-to-miss

[8] Va sottolineato, in coerenza, anche il fenomeno del "disimpegno morale" (Bandura, 1999) quale insieme di meccanismi attraverso i quali gli individui razionalizzano comportamenti eticamente e socialmente problematici.

[9] Si veda: https://joint-research-centre.ec.europa.eu/scientific-activities-z/survey-methods-and-analysis-centre-smac/loneliness/loneliness-prevalence-eu_en; Schnepf *et al.*, 2024. https://link.springer.com/book/10.1007/978-3-031-66582-0#about-authors.

L'IA, in questo quadro, si applica a nuove forme di relazione: a seconda della piattaforma e del dispositivo tecnologico, queste relazioni tecnologiche possono ridurre la solitudine o, al contrario, aumentare l'alienazione.

Un esempio significativo è rappresentato dai *Lovot*, "robot emotivi" progettati dalla giapponese *Next Robotics* per offrire compagnia e migliorare il benessere socio-psicologico. A differenza degli assistenti virtuali, i *Lovot* non svolgono funzioni pratiche, ma sono ideati per interagire affettuosamente con gli utenti, apprendendo dalle interazioni grazie all'IA. Reagiscono al tocco, ai suoni e alle espressioni facciali, mostrando comportamenti empatici che favoriscono un legame emotivo, simile a quello con un animale domestico (Zhou, Fischer, 2019).

Un'altra applicazione è *Replika-My AI Friend***,** un chatbot basato su IA che simula conversazioni per offrire compagnia e supporto emotivo. Grazie all'apprendimento automatico, Replika sviluppa una personalità unica, adattandosi alle emozioni e preferenze dell'utente, che può considerarlo un amico, un confidente o persino un partner romantico. Sebbene molti utenti trovino conforto in tali interazioni[10] (Pentina *et al*., 2023), esiste il rischio di sviluppare una dipendenza emotiva che può aggravare il senso di isolamento, alterare la percezione della realtà ed esacerbare condizioni preesistenti come ansia e depressione (Laestadius *et al.*, 2024).

4. Tra sfide e potenzialità dell'IA come "agente relazionale"

L'applicazione quotidiana e globale dell'IA – anche nella sfera relazionale, affettiva ed emotiva – porta con sé nuovi interrogativi che necessitano di un approccio sociologico critico ed integrato.

Tradizionalmente, l'IA è stata vista come un insieme di algoritmi e tecniche computazionali progettate per risolvere problemi specifici; tuttavia, essa ha cominciato a giocare un ruolo più dinamico nelle relazioni interpersonali, sviluppando capacità di comunicazione e di adattamento emotivo che le permettono di "relazionarsi" con gli esseri umani.

Un "agente relazionale" è un'entità che non solo esegue compiti e fornisce assistenza, ma che può anche stabilire e mantenere rapporti con gli individui, rispondendo alle loro emozioni, creando interazioni che simulano quelle interpersonali, influenzando i loro comportamenti.

[10] Sul tema: Sheng A., Wang, F. (2022).

L'IA come "agente relazionale" rappresenta, allora, una nuova frontiera che offre sia opportunità sia sfide significative: ha il potenziale di migliorare il benessere umano (Fitzpatrick *et al.*, 2017), è capace di personalizzare le esperienze e può "alleviare" l'esperienza di isolamento.

Parallelamente, solleva questioni di devianza emozionale, alienazione, manipolazione, solitudine che potrebbero compromettere la qualità delle relazioni insieme all'autonomia degli individui.

Siamo di fronte ad una trasformazione sistemica che investe il Welfare e l'intero sistema socio-assistenziale, ridefinendo le modalità di presa in carico e di progettazione degli interventi a sostegno delle vulnerabilità individuali e gruppali in un'ottica olistica, ma al contempo rischiando di amplificare le disuguaglianze e le fragilità sociali esistenti.

Conclusioni

La sociologia contemporanea non si limita ad esplorare le potenzialità dell'IA come agente relazionale, ma si configura sempre più come campo di indagine critico orientato a decifrare le trasformazioni strutturali dei regimi affettivi nella società onlife (es. la solitudine). In tale contesto, l'affettività e le emozioni non possono essere più concepite esclusivamente come dimensioni private, ma vanno analizzate come un costrutto socio-tecnologico, profondamente riconfigurate dalle logiche algoritmiche, dalle piattaforme digitali e dai dispositivi di interazione mediata (Illouz, 2007; Lupton, 2014). Le cosiddette "tecnologie affettive" pongono, in altri termini, nuove sfide interpretative alla comprensione sociologica della relazione, della cura e della dipendenza intersoggettiva. Risulta, pertanto, centrale indagare come tali tecnologie plasmino le dinamiche del riconoscimento sociale e l'organizzazione collettiva dell'empatia, fino a ridefinire i confini di ciò che viene considerato legittimamente "umano". In questa prospettiva, diventa urgente elaborare una rinnovata direzione di ricerca sociologica che, in un'ottica critica e pubblica, sia in grado di interrogare i processi attraverso cui affettività ed emozioni vengono codificate, mediate e istituzionalizzate nei contesti onlife.Tale approccio dovrebbe includere, da un lato, la promozione di un'educazione socio-emotiva onlife, intesa come pratica formativa integrata che riconosca la co-presenza di dimensioni corporee, digitali e relazionali; dall'altro, l'estensione del concetto di "lavoro emotivo" (Hochschild, 1983), includendo le interazioni quotidiane mediate da dispositivi tecnologici e sistemi di IA, con particolare attenzione agli aspetti normativi, simbolici e performativi di tali pratiche affettive. Soprattutto

da parte delle giovani generazioni. È chiaro che, alla luce di quanto detto, la distinzione tra umano e artificiale non può più essere assunta come fissa o naturale, ma va problematizzata come costruzione sociale in continua negoziazione, densa di implicazioni epistemologiche, etiche, tecnologiche e politiche. In tal senso, l'urgenza teorica ed empirica di una sociologia delle emozioni si configura non solo come strumento analitico per interpretare le trasformazioni sociali in atto, ma anche come presa di posizione critica articolata su più livelli (micro, meso e macro) nella ridefinizione dell'esperienza socio-relazionale contemporanea.

Riferimenti bibliografici

Andries V., Robertson J. (2023). Alexa doesn't have that many feelings: Children's understanding of AI through interactions with smart speakers in their homes. *Computers and Education: Artificial Intelligence*, 5: 100176. https://doi.org/10.1016/j.caeai.2023.100176

Bandura A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, 3(3): 193-209.

Bauman Z. (1999). *Modernità liquida*. Roma-Bari: Laterza, 2011.

Beck A., Giddens A., Lash S., et al. (1994). *Modernizzazione riflessiva. Politica, tradizione ed estetica nell'ordine sociale della modernità*. Trieste: Asterios, 1999.

Butler J. (1990). *Gender trouble: feminism and the subversion of identity*. London: Routledge.

Cambi F. (1998). *Nel conflitto delle emozioni*. Roma: Armando.

Castells M. (2002). *La nascita della società in rete*. Milano: Egea, 1996.

Cerulo M. (2009). *Il sentire controverso. Introduzione alla sociologia delle emozioni*. Roma: Carocci.

Cerulo M. (2014). *La società delle emozioni. Teorie e studi di caso tra politica e sfera pubblica*. Napoli-Salerno: Orthotes.

Cerulo M. (2018). *Sociologia delle emozioni*. Bologna: il Mulino.

Cerulo M. (2019). Fare società attraverso le emozioni. Un'analisi di alcune forme di agire sociale affettivo. *Sociologia*, 3: 57-62.

Collins R. (1981). On the microfoundations of macrosociology. *American Journal of Sociology*, 86(5): 984-1014.

Collins R. (2014). Interaction ritual chains and collective effervescence. In Von Scheve C., Salmela M. (eds.), *Collective emotions. Perspectives from psychology, philosophy and sociology*. Oxford: Oxford University Press.

Couldry N., Hepp A. (2017). *The mediated construction of reality*. Cambridge: Polity Press.

D'Ambrosio M. (2019). La sociologia delle emozioni e il legame sociale. Ripensare il rapporto "tra umani" nella società tecnologica. *Culture e Studi del Sociale*, 4(2): 177-192. http://www.cussoc.it/index.php/journal/issue/archive

D'Ambrosio M., Barba D. (2023). Il bisogno affettivo e l'inganno dei social: i presupposti e le pratiche del Romance Scam. *Rivista di Criminologia, Vittimologia e Sicurezza*, XVII: 74-87.

Donath J. (2014). *The social machine: designs for living online*. Cambridge (MA): MIT Press.

Durkheim É. (1895). *Le regole del metodo sociologico*. Milano: Edizioni di comunità, 1963.

Elias N. (1987). *La società di corte*. Bologna: il Mulino.

Feenberg A. (2002). *Tecnologia in discussione. Filosofia e politica nella moderna società tecnologica*. Milano: Etas.

Fiske A., Henningsen P., Buyx A. (2019). Your robot therapist will see you now: ethical implications of embodied artificial intelligence. *Journal of Medical Ethics*, 45(5): 317-324.

Fitzpatrick K., Darcy J.R., Vierhile J.E. (2017). Delivering cognitive behavioral therapy to adolescents with anxiety using an automated chatbot: a randomized controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, 56(10): 857-864.

Floridi L. (2014). *The onlife manifesto. Being human in a hyperconnected era*. London: Springer, 2015.

Guzman A.L. (2018). Imagining the voice in the machine: the ontology of digital social agents. *Media, Culture & Society*, 40(1): 62-76.

Hochschild A. (1983). *The managed heart: commercialization of human feeling*. Berkeley (CA): University of California Press.

Hyniewska S., Sato W. (2015). Facial feedback affects valence judgments of dynamic and static emotional expressions. *Frontiers in Psychology*, 6: 291. https://doi.org/10.3389/fpsyg.2015.00291

Keysers C., Kaas J.H., Gazzola V. (2010). Somatosensation in social perception. *Nature Reviews Neuroscience*, 11(6): 417-428. https://doi.org/10.1038/nrn2833

Kundi B., El Morr C., Gorman R., Dua E. (2023). Artificial intelligence and bias: a scoping review. In El Morr C. (2022). *AI and society: tensions and opportunities* (pp. 199-215). Boca Raton: Chapman and Hall/CRC. https://doi.org/10.1201/9781003261247

Kurian N. (2023). AI's empathy gap: the risks of conversational artificial intelligence for young children's well-being and key ethical considerations for early childhood education and care. *Contemporary Issues in Early Childhood*, 14639491231206004. https://doi.org/10.1177/14639491231206004

Kurian N. (2024). 'No, Alexa, no!': designing child-safe AI and protecting children from the risks of the 'empathy gap' in large language models. *Learning, Media and Technology*: 1-14. https://doi.org/10.1080/17439884.2024.2367052

Laestadius L., Bishop A., Gonzalez M., Illenčík D., Campos-Castillo C. (2024). Too human and not human enough: a grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. *New Media & Society*, 26(10): 5923-5941. https://doi.org/10.1177/14614448221142007

Lupton D. (2014). *Digital sociology*. London-New York: Routledge.

Marchetti A., Massaro D. (2023). *Robot sociali e educazione. Interazioni, applicazioni e nuove frontiere*. Milano: Raffaello Cortina Editore.

Pajno A. et al. (2019). AI: profili giuridici. Intelligenza artificiale: criticità emergenti e sfide per il giurista. *BioLaw Journal*, 3: 205-235.

Pentina I., Hancock T., Xie T. (2023). Exploring relationship development with social chatbots: a mixed-method study of Replika. *Computers in Human Behavior*, 140: 107600.

Ricucci R., Sannella A. (a cura di) (2024). *Dizionario di sociologia per la persona*. Milano: FrancoAngeli.

Santamaria M. (2019). *Tinder and the city. Avventure e disagi nel mondo delle dating app*. Milano: Agenzia Alcatraz.

Schnepf et al. (2024). *Loneliness in Europe. Determinants, risks and interventions*. Cham: Springer. https://link.springer.com/book/10.1007/978-3-031-66582-0#about-authors

Sheng A., Wang, F. (2022). Falling in love with machine: Emotive potentials between human and robots in science fiction and Reality. *Neohelicon,* 49(2): 563–577.

Thoits P.A. (1985). Self-labeling processes in mental illness: the role of emotional deviance. *American Journal of Sociology*, 91(2): 221-249. https://doi.org/10.1086/228276

Thoits P.A. (1990). Emotional deviance: research agendas. In Kemper T.D. (ed.), *Research agendas in the sociology of emotions* (pp. 180-203). Albany (NY): State University of New York Press.

Turkle S. (1995). *Life on the screen: identity in the age of the internet*. New York: Simon & Schuster.

Turkle S. (2011). *Alone together: why we expect more from technology and less from each other*. New York: Basic Books.

Turnaturi G. (1995). *La sociologia delle emozioni*. Milano: Anabasi.

Weber M. (1922). *Economia e società* (tr. it. 2005). Roma: Donzelli.

Zhou Y., Fischer M.H. (2019). *AI love you: developments in human-robot intimate relationships*. Cham: Springer.

Zuboff S. (2019). *The age of surveillance capitalism: the fight for a human future at the new frontier of power*. New York: PublicAffairs.

La religione dell'intelligenza artificiale: la Babele dei Santi, Patroni e divinità

di *Maria Chiara Spagnolo*[*]

La scoperta dell'intelligenza artificiale ha reso possibile incarnare l'elemento umano nella mente immortale della macchina rendendola divina. In queste circostanze può sorgere la tentazione di creare un *by-pass* religioso rispetto all'attuale condizione dell'uomo, che rischierebbe di chiudersi nel mondo dell'Intelligenza artificiale e dei nuovi motori di ricerca che replicano antiche fantasie religiose trasformandole in miti contemporanei.

Parole chiave: religione; intelligenza artificiale; santi; motori di ricerca; miti; sacro.

The religion of artificial intelligence: the Babel of Saints, Patrons and Gods
The discovery of artificial intelligence has made it possible to embody the human element in the immortal mind of the machine, making it divine. In these circumstances, the temptation may arise to create a religious by-pass with respect to the current condition of man, who would risk closing himself off in the world of Artificial Intelligence and the new search engines that replicate ancient religious fantasies by transforming them into contemporary myths.

Keywords: religion; artificial intelligence; saints; search engines; myths; sacred.

Introduzione

Nella società contemporanea la tecnica ha consentito di sottrarre il dolore dall'ordinario della vita e ha così permesso di coltivare l'illusione che non vi sia più sofferenza o che essa sia qualcosa di neutralizzabile. Ma la neutralizzazione del dolore non ne annulla l'esistenza.

Sorge, così, un incoercibile bisogno di salvezza intramondana che pretende un mondo senza dolore, di conseguenza la società ha prodotto una sorta di neopaganesimo e un bisogno di salvezza senza fede tenuto insieme dalla tecnica. Nella tecnica si è oggettivata quel che un tempo si chiamava anima o mente, mentre la ragione è divenuta una potenza impersonale e manipolatrice (Collins, 2018).

DOI: 10.5281/zenodo.17524092
[*] Università del Salento. mariachiara.spagnolo@unisalento.it.

In questo senso, la tecnica tende a spiritualizzare i corpi senza che sia necessaria la resurrezione della carne. Il progresso tecnologico ha attenuato il peso del dolore, ma a fronte di un tale vantaggio si disegnano anche nuovi limiti. Un limite singolare è dato dal fatto che nella società contemporanea non si sa più o si sa sempre meno, come allocare il dolore nella vita, indipendentemente dalla capacità che i soggetti hanno di adattarvisi e di comprenderlo (Stahl, 1999). Nel cristianesimo il dolore era in qualche modo giustificato e reso riscattabile, al contrario, nella società attuale le intensificate possibilità di vita rendono sempre meno concepibile il dolore occultandolo o raggirandolo.

La scoperta dell'intelligenza artificiale ha reso possibile incarnare l'elemento umano nella mente immortale della macchina – rendendola divina. Dalle macchine di Turing e dal 1936 in poi, non solo il corpo umano è stato trasceso, ma con esso, anche l'intelligenza, fino ad arrivare alla costruzione di macchine autonome e dotate di intelligenza capaci di sostituire la controparte umana.

Tuttavia, la macchina formale, ideata da Turing si presentava solo come un sistema astratto che, se adeguatamente programmato, era capace di eseguire ogni tipo di operazione, ma non era – perlomeno non ancora – ciò che i futuri studiosi avrebbero chiamato o paragonato ad un "nuovo testamento dell'era digitale" (Dyson, 2012) e dell'informatica attuale. Turing sicuramente era il crinale tra la de-corporalizzazione di Cartesio, la logica di Leibniz e le macchine da guerra di Von Neumann, nelle cui teorie e ricerche la mente pensante è soprattutto misurante, nella costituzione di un sapere ontologico che dall'uomo si incarna nei mezzi e nelle macchine. È questo il passaggio che precede le congetture di Turing, l'anima, il pensiero umano che trasmigra come perfezione dall'Ente supremo, all'uomo, alle macchine, con la possibilità dell'implementazione del sapere: la macchina dimostra la macchina proprio partendo da sé stessa, esattamente come l'esistenza di Dio si dimostra a partire dallo stesso concetto di Dio, senza alcuna contraddizione, ma incamerando la grandezza come suo attributo.

La tesi di Turing si basa proprio sull'idea che, se esiste un algoritmo per eseguire dei compiti e manipolare dei simboli, allora esiste una macchina in grado di eseguire quel compito. Di conseguenza, è possibile utilizzare il modello della macchina di Turing per definire i limiti della computabilità: se per un problema non esiste una macchina di Turing in grado di risolverlo allora il problema si dice incomputabile o irrisolvibile. In *On Computable Numbers* del 1936, la macchina di calcolo logico non è presentata come una macchina fisica, ma come un "ideal-tipo", un modello ideale e intelligente al quale rifarsi, un modello in cui tempo e spazio appaiono come infiniti: una sintesi tra religione e tecnologia.

Negli anni Ottanta, soprattutto nelle visioni dei ricercatori dell'ARPA (Advanced Research Projects Agency) e in ambiente militare, si era concretizzata l'dea di un "trasferimento" della mente in una macchina, di una mente umana in

una "rete neurale artificiale", creando a tutti gli effetti un'anima in un universo sempre più artificiale e promuovendo ciò che Noble chiama "religione della tecnologia" (1997).

In queste circostanze può sorgere – nell'ambito delle religioni – la tentazione di creare un *by-pass* alla condizione odierna dell'uomo nel mondo, che rischierebbe di chiudersi in un universo altro dotato di un proprio senso religioso (Campbell, 2010; Ellul 1964).

Il sito Prega.org si presenta come una variante devota di ChatGpt uno spazio di intelligenza artificiale in cui chattare con vari santi e con Padre Pio. Simile, l'esperimento attuato durante l'annuale *Kirchentag* della Chiesa evangelica tedesca, che ha affidato a ChatGpt – proiettando su un grande schermo l'immagine di un pastore virtuale –, la proclamazione del sermone e l'organizzazione dell'intero culto (canti, preghiere). Ancora, l'app *Text with Jesus* lanciata sul mercato come piattaforma di messaggistica istantanea, promette un collegamento ipertestuale con Gesù (Defamer, 2007).

Hozana, invece è un social network di preghiere che propone delle linee guida su come pregare, oltre a elencare delle differenze tra le varie richieste di preghiere (*lamentatio*, dono, conforto). Ancora una volta, nel sito appare evidente come la mancanza di tempo sia un problema della modernità. La preghiera *fast* richiede un minuto, la promessa di una spiritualità immediata sembra essere soddisfatta dall'ampia scelta del *bric à brac* religioso. Le macchine, perciò, disciplinano e dettano le regole delle relazioni tra gli attori sociali anche a livello dell'esperienza religiosa (Bingaman, 2020).

Il contributo vuole dimostrare come l'intelligenza artificiale offra la possibilità di replicare antiche fantasie religiose trasformandole in miti contemporanei, dove tutto è divinizzato e anche Maria può interpretare il ruolo di una fervente femminista. Ci si pone il problema se sarà ancora più difficile credere, trasformare il dolore in nuovi culti o se le religioni parleranno la stessa lingua, superando l'antico mito di Babele.

1. Ciò che la tecnologia promette in termini magici e poi religiosi

"Chi vuole avere le visioni vada al cinematofrago", così sinteticamente sentenziava Max Weber in *Considerazioni Intermedie* collegando le nuove tecnologie della comunicazione al linguaggio religioso. Le discussioni sociologiche all'interno dell'avvento della prima modernità hanno spesso affrontato la nascita della tecnologia come una improvvisa spinta, una "tensione" valoriale e dei tempi che attraverso il linguaggio (anche spirituale) ha provato a descrivere il cambiamento del potere, del capitale, del controllo e l'efficienza esponenziale e

incontrollabile della tecnologia (Ellul, 1964; Nye, 2004). Studiosi come Weber, Sombart, Simmel, Marx, Benjamin, che nei loro lavori e analisi alludevano alle promesse e ai pericoli della tecnologia con toni religiosi, l'hanno presentata – in alcuni casi – come una forza potente che libera (carisma) o che controlla gli attori sociali (massificazione delle merci, dell'arte, reificazione). Pertanto, lo studio della tecnologia è stato strettamente collegato all'uso del linguaggio, alle metafore o alle immagini religiose che esso crea e che descrivono in modo efficace le relazioni tra uomo e tecnica.

La magia della parola e del guardare, la magia dell'inaspettato e dell'invenzione che sin dall'avvento della stampa con la sua divulgazione della protesta religiosa ha identificato e ha agganciato la tecnologia al sistema religioso (Florenskij, 2023; Durham, 2005).

La religione come comunicazione mediata dal computer è diventata un consolidato oggetto di studio negli ultimi venti anni. Internet ha dato voce e ha fatto emergere una pluralità significativa di attori religiosi che hanno ben compreso le potenzialità dello strumento, utilizzato per allargare e amplificare il raggio della loro azione comunicativa. Quasi nulla di nuovo rispetto al passato, quando le grandi chiese o i predicatori indipendenti hanno utilizzato i mezzi di comunicazione di massa (radio, televisione) per entrare nelle case e richiamare l'attenzione della gente che non frequentava più i luoghi di culto.

Tra il 1970 e il 1990 negli Stati d'Uniti d'America, pastori protestanti hanno fondato chiese personali, gestendo direttamente reti televisive, predicando la Bibbia con uno stile da *talkshow* e televendita.

Allo stesso tempo, negli ultimi anni, si sono moltiplicati luoghi o spazi elettronici che hanno ospitato prodotti delle più svariate chiese, confessioni o gruppi religiosi che attraverso i mezzi tecnologici hanno promosso e fatto circolare idee, messaggi etici e teologici, e predicazioni del 'buon vivere'. Le *Electronic Churches* o *Internet Churches* sono le capostipiti della fiducia costruita a distanza tra il mezzo elettronico e il fedele, in cui il corpo si rivela come la minima parte di quel processo fondativo ed elaborativo del fenomeno di espropriazione della religione dal suo leader carismatico come viatico (Pace, 2021).

Con l'elaborazione dei computer, dei cellulari e tablet, così come l'avvento di Internet, la relazione tra l'impegno fideistico umano e l'avanzamento tecnologico ha prodotto nuove metafore comunicativo/religiose che si sono ulteriormente rafforzata con l'avvento dell'intelligenza artificiale.

È diventato comune usare miti che collegano le tecnologie create dall'uomo a uno scopo o risultato superiore e trascendente. I miti forniscono strumenti per interpretare la complessa relazione degli esseri umani con i loro strumenti tecnologici. Alcuni di questi miti forniscono utili spunti riguardo la cornice spirituale che collega l'essere umano alla tecnologia. Noble, nel suo lavoro sulla

storia dell'impresa tecnologica, presenta il mito della "religione della tecnologia", in base al quale la tensione umana verso la progressione tecnologica è un tentativo di riconquistare il senso perduto di divinità: anche quello della mente umana come "perfezione divina" (Noble, 1999). La "techgnosis" (Davis, 2023), come mito e ossessione del mondo occidentale per la tecnologia ha conferito all'uomo dei poteri idealizzati e percepiti come divini sin dalle origini e dalle prime scoperte (oggi Google e le industrie informatiche e tecnologiche della Silicon Valley). La tecnologia è antropomorfizzata, essa catalizza le aspettative future ed esperienziali dell'esistenza umana attraverso ciò che Davis chiama la "tecnomistica" della società. La tecnologia è dunque una forza che spinge l'umanità, una forza che appare incontrollabile quando è percepita come un'entità a sé e non più sviluppata dall'uomo stesso. Ed è qui che si radica la fede universale nella tecnologia e che costruisce un sistema di credenze simili all'esperienza mistico/religiosa, con simboli e rituali anche al di fuori delle stesse organizzazioni religiose ufficiali.

Una panoramica di questi miti evidenzia tre distintive narrazioni di inquadramento sulla correlazione tra religione e tecnologia: a) la tecnologia offre la redenzione umana e l'umanità in maniera collettiva diventa simile a Dio abbracciando la tecnologia; b) la tecnologia stessa è una forza divina o spirituale; c) la tecnologia offre agli esseri umani un'esperienza magica o religiosa al di fuori delle istituzioni e degli ambienti tradizionali.

Tuttavia, l'intersezione tra tecnologia e metafisico spesso provoca un senso di smarrimento o di angoscia rispetto alle risposte che in alcuni casi risultano standardizzate. Perciò questi miti sollevano anche tre potenziali aree di vuoto:

1) circa la complessa natura dell'umanità; 2) la natura della tecnologia e i suoi risvolti etico/morali; 3) la relazione dell'uomo come inventore e consumatore delle sue stesse scoperte tecnologiche. Alcuni miti nati dall'interazione della tecnologia e dal suo consumo – più che mero uso – suggeriscono che il simbolismo religioso può essere facilmente importato nelle discussioni riguardanti lo sviluppo della tecnologia all'interno della società.

Basti pensare alla nascita del Mac come mito della creazione e della conoscenza attraverso il simbolo della mela, ma anche come rottura ed evento trasformativo di un'intera cultura: ciò che non esisteva prima e ciò che sarà dopo.

Dalla fine degli anni Novanta ad oggi l'attenzione degli studiosi interessati al tema delle religioni nel web si è sempre più spostata dalla TV ad internet, fino a giungere all'intelligenza artificiale. La distinzione adottata da Christopher Helland fra *religion online* –istituzioni religiose che si adattano a comunicare via internet – e *online religion*, nuovi network capaci di promuovere la formazione di comunità religiose virtuali nelle quali la definizione dei contenuti e dei significati religiosi o spirituali è affidata all'interazione via computer fra gli individui

– appare svuotata di senso, così come lo sono i termini "virtuale" e "cyberspazio".

Il cambiamento socioculturale che si è determinato negli ultimi anni e poi amplificato dall'introduzione dell'IA, pone un netto distacco tra le religioni fai da te e le religioni in cui la stessa intelligenza artificiale appare come l'unico mediatore culturale e religioso tra Dio e gli uomini.

In questo caso, la religione e il suo messaggio non sono più mediati, veicolati dal mezzo, dall'*usenet*, dal *blog* o da qualsiasi canale social, ma da un'Intelligenza, da un Altro Motore spirituale, a cui rivolgere le proprie perplessità. La profezia dell'etica rousseauniana «quanti uomini tra Dio e me» si avvera nell'abolizione dei corpi e nella trascendenza della macchina. Ma chi, in questo caso, garantisce l'autorità del messaggio?

2. Dialoghi artificiali

Sebbene l'intelligenza artificiale non possa sperimentare su sé stessa il "sentire" dell'esperienza religiosa o spirituale, può però potenziare le pratiche religiose, così come le forme comunicative, la ricerca dell'ascolto e le nuove forme di spiritualità che da essa derivano.

L'esperienza religiosa è più accessibile, immersiva e interattiva, le domande appaiono come naturali e soprattutto con la nascita esponenziale di *app* religiose che simulano attraverso l'IA il dialogo con i Santi, la paura del peccato e del giudizio individuale appare più sfumato.

Tuttavia, resta da capire come queste innovazioni influenzeranno il significato e l'autenticità della religione, soprattutto in contesti più tradizionali e ufficiali.

L'uso dell'intelligenza artificiale per simulare esperienze religiose più profonde e creare comunità virtuali o messe virtuali è un campo emergente, ma ci sono già alcuni esempi concreti che esplorano possibilità in cui la partecipazione a cerimonie religiose in ambienti virtuali risulta immersiva. The *Temple of the Mind* è un progetto di realtà virtuale e intelligenza artificiale che personalizza le esperienze basate sui dati cognitivi dell'utente (elaborazione delle emozioni e stati mentali). Il progetto mira a generare ambienti virtuali che stimolano sensazioni di trascendenza simili a quelle che alcuni soggetti vivono durante le esperienze estatico/religiose. Questi ambienti virtuali possono includere paesaggi mistici, suoni rilassanti o meditativi e simbolismi religiosi che guidano l'utente e lo incoraggiano nell'avere una "comunicazioni con il divino".

Piattaforme come Second Life e VRChat, attirano utenti di tutto il mondo e che a loro volta creano comunità virtuali per partecipare ad eventi religiosi,

celebrare particolari festività o discutere di tematiche spirituali. Sebbene queste piattaforme non siano specificamente religiose, molte chiese e gruppi spirituali hanno creato spazi virtuali dove i membri possono partecipare a cerimonie religiose con preghiere di gruppo o meditazioni collettive.

L'intelligenza artificiale in questo contesto è usata per moderare le discussioni, creare eventi personalizzati e sviluppare ambienti virtuali che rappresentano dei veri e propri luoghi del sacro. Anche in Minecraft alcuni utenti hanno creato chiese virtuali o luoghi di culto, dove le persone possono incontrarsi e pregare insieme. In queste comunità, l'IA facilita gli incontri, suggerisce letture religiose e guida preghiere collettive. Sebbene non si tratti a tutti gli effetti di esperienze mistico-spirituali, queste piattaforme offrono una forma di "community building" spirituale che fa uso di tecnologie moderne.

Stessa cosa accade per l'elaborazione di messe virtuali e servizi religiosi che si servono dell'IA per migliorare i loro contenuti. In questo caso, l'intelligenza virtuale è usata per analizzare le preferenze individuali dei partecipanti (tipo di musica, predica, preghiera) e suggerire contenuti religiosi differenziati per soggetto.

Alcuni strumenti di IA potrebbero persino creare prediche su misura, analizzando i temi trattati nelle liturgie precedenti e producendo sermoni che rispondano alle domande e ai bisogni spirituali della comunità. Un altro campo di azione è la produzione di assistenti virtuali i quali rispondono a domande religiose, offrono preghiere personalizzate, o guidano l'utente durante la messa virtuale. Un esempio è "Reverend AI", un assistente religioso che può predicare, rispondere a domande bibliche, o avviare preghiere durante una messa virtuale. L'uso dell'IA in questo contesto mira a rendere i servizi religiosi più accessibili, specialmente per chi ha difficoltà a partecipare fisicamente o preferisce un'esperienza più intima e personalizzata.

In Italia, al momento non ci sono esempi noti di Reverend AI, ma in alcuni casi, le app che si basano sull'utilizzo dell'intelligenza artificiale servono per creare delle preghiere individualizzate, proporre pratiche di meditazione, suggerire letture bibliche secondo le preferenze del consumatore religioso.

In Europa, l'uso di assistenti virtuali religiosi basati sull'intelligenza artificiale è ancora in una fase sperimentale.

In Germania, sono stati avviati alcuni esperimenti di robotica religiosa, come il progetto che ha coinvolto un robot sacerdote. Il robot, chiamato "BlessU-2", è stato costruito da un gruppo di ingegneri e teologi e offre benedizioni ai fedeli. Il robot è in grado di benedire, alzare le mani, ma non è dotato di un'intelligenza artificiale autonoma che genera sermoni e riflessioni teologiche. Sempre in Germania, durante l'annuale *Kirchentag* della Chiesa evangelica tedesca (EKD), è stato utilizzato ChatGPT per predicare durante un evento pubblico, proiettando

sul grande schermo l'immagine di un pastore virtuale che ha guidato i partecipanti nella riflessione su temi religiosi.

Nel Regno Unito il chatbot "Clergy-bot" risponde a domande religiose, pur non essendo ancora in grado di condurre cerimonie e prediche complesse.

In Finlandia, gli sviluppatori del Progetto "A.I Religion" hanno creato algoritmi in grado di produrre considerazioni sulla religione, citazioni bibliche e meditazioni spirituali, adattandole alle esigenze degli utenti.

In Giappone, all'interno di un tempio Buddista esiste un robot chiamato Mindar che risponde alle preghiere dei fedeli.

Hozana invece, è una piattaforma digitale cattolica che offre uno spazio online per la preghiera e la spiritualità. Fondata nel 2016, Hozana mira a creare una comunità di preghiera globale, dove le persone possono trovarsi e partecipare a iniziative religiose a distanza.

La piattaforma offre la possibilità di ricevere preghiere quotidiane via e-mail, che possono essere tematiche o liturgiche. Gli utenti possono scegliere di ricevere preghiere specifiche, come quelle per la giornata, per determinati eventi religiosi o anche per meditazioni personali.

La piattaforma collabora con molte istituzioni religiose, come parrocchie, ordini monastici, associazioni cattoliche e scuole di spiritualità; perciò, Hozana si propone come una piattaforma inclusiva, aperta a chiunque voglia approfondire la propria fede attraverso la preghiera online, senza distinzione di nazionalità, lingua o appartenenza religiosa.

Conclusioni

Le tecnologie avanzate, potrebbero evocare un senso di meraviglia e misticismo simile all'esperienza religiosa. L'idealizzazione dell'IA porta l'utente ad immaginare una religione basata sulla tecnologia molto più semplificata, spogliata da dogmi e che può risolvere problemi non solo di ordine spirituale, ma anche di ordine etico/morale che vanno oltre l'attanagliamento del giudizio e dell'ammenda religiosa.

In un mondo sempre più disinteressato alle tradizionali narrazioni religiose l'intelligenza artificiale potrebbe essere vista come una sorta di sostituto della fede, non un surrogato, ma un dio del bisogno immediato. L'uso dell'IA non pone delle riflessioni sul futuro della fede – troppe sono state le domande e le analisi condotte sulle trasformazioni socio-culturali delle "vie" del sacro e del suo *camouflage* – ma su come la tecnologia possa offrire delle possibilità nella crescita spirituale, nelle pratiche e culti di fede.

Non si tratta di rendere le religioni più accessibili o esemplificarle, scarnificarle o disumanizzarle, ma di come invece, il sacro sia ancora inteso come una categoria imprescindibile che accende e anima l'"effervescenza" sociale.

Di fatto, esperimenti di IA tramite l'utilizzo di robot o immagini proiettate che celebrano culti e rispondono alle domande più disparate, sono la dimostrazione di quanto ancora il sacro abbia bisogno di una forma incarnata nel materiale. Il robot non fa altro che sostituire il muto santo avvolto nella sua ceroplastica e la statua come *plena deo*, ma anch'essa silente. In un mondo in cui il «brusio» non è più neppure avvertito, ci si rifugia in dialoghi (più che pratiche) singolari, senza chiedersi che cosa sia giusto o sbagliato, ma alterando, semmai, le dinamiche del potere e del controllo rispetto alla tradizione: quello religioso e quello del capitale.

Nei romanzi e nelle storie di fantascienza la religione e le religioni altre, salvano gli individui e i piccoli gruppi scelti dall'andazzo di "questo" mondo. Pochi eletti. Per chi ci crede, per chi segue.

Dove ci porterà l'IA, in quale mondo? Quale destinazione?

L'idea di un mondo religioso promesso dall'intelligenza artificiale è un argomento affascinante e complesso, che può essere interpretato in molti modi diversi, purché essa stessa non si considerata come una religione: pena il divenire poi obsoleta e poco accattivane.

Riferimenti bibliografici

Campbell H. (2010). *When Religion Meets New Media*. London: Routledge. https://doi.org/10.4324/9780203695371

Collins H. (2018). Artifictional intelligence: against humanity's surrender to computers. AI & SOCIETY (2019): Cambridge. UK.

Davis E. (1998). *Techgnosis*. New York: Harmony Books.

Davis E. (2023). *Techgnosis. Mito magia e misticismo nell'era dell'informazione*. Roma: Nero.

Defamer (2007). Short ends: The Jesus Phone finally arrives. 9 January. http://defamer.com/hollywood/short-ends/short-ends-the-jesus-phone-finally-arrives-227604.php

Durham P.J. (2005). *Parlare al vento. Storia dell'idea di comunicazione*. Milano: Booklet.

Dyson G. (2012). *La cattedrale di Turing*. Torino: Codice Edizioni.

Ellul J. (1964). *The Technological Society*. New York: Alfred A. Knopf.

Florenskij P. (2003). *Il valore magico della parola*. Milano: Medusa.

Noble D.F. (1999). *La religione della tecnologia. Divinità dell'uomo e spirito d'invenzione*. Torino: Edizioni di Comunità.

Pace E. (2021). *Introduzione alla sociologia delle religioni*. Roma: Carocci.

Wolf A. (1991). Mind, self, society, and computer: Artificial intelligence and the sociology of mind. *American Journal of Sociology*, 96(5): 1073-1096. https://doi.org/10.1086/229649

Responsabilità e implicazioni etiche dei Sistemi GPS d'emergenza. Il caso dell'eCall e l'integrazione dell'IA nei veicoli

di *Michela Morelli*[*]

Il malfunzionamento di sistemi GPS d'emergenza, come l'eCall, evidenzia criticità e rischi etici. Errori automatizzati possono configurare reati come procurato allarme e generare responsabilità civile, rendendo urgente chiarire competenze e valutare i rischi per la sicurezza. L'articolo analizza il quadro normativo ed etico, mettendo in luce i principali rischi connessi a tali errori.

Parole chiave: malfunzionamento sistemi informatici; responsabilità civile, responsabilità penale; intelligenza artificiale; veicoli; implicazioni etiche.

Responsibilities and ethical implications of emergency GPS systems. The case of eCall and the integration of AI into vehicles

The malfunction of emergency GPS systems, such as eCall, highlights critical issues and ethical risks. Automated errors may give rise to offenses such as false alarm and may also entail civil liability, thereby underscoring the urgency of clarifying responsibilities and assessing safety risks. This article examines the regulatory and ethical framework, shedding light on the main risks associated with such malfunctions.

Keywords: malfunctioning of computer systems; civil liability, criminal liability; artificial intelligence; vehicles; ethical implications.

Introduzione

Il tema dell'obbligatorietà dei sistemi di emergenza nei veicoli si colloca in un punto centrale del dibattito su progresso tecnologico, sicurezza collettiva, diritto ed etica. L'obbligo sancito dal Regolamento UE 2015/758, che ordina l'installazione del sistema eCall, rappresenta una svolta: si stima una riduzione dei tempi di intervento fino al 50% e circa 2.500 vite salvate ogni anno, con benefici economici di 26 miliardi di euro.

DOI: 10.5281/zenodo.17524133

[*] Università degli studi del Molise. m.morelli3@studenti.unimol.it.

L'adozione di tali dispositivi pone interrogativi che vanno oltre l'efficienza funzionale. Le implicazioni giuridiche, etiche ed infrastrutturali connesse all'uso diffuso di tecnologie automatizzate – specie se integrate con intelligenza artificiale – impongono una riflessione necessariamente interdisciplinare. Non basta valutare l'affidabilità tecnica dei sistemi, ma occorre considerare la distribuzione delle responsabilità in caso di malfunzionamenti, la trasparenza e l'intelligibilità degli algoritmi decisionali e la compatibilità dell'innovazione digitale con la tutela dei diritti fondamentali.

Alla luce di ciò, il presente contributo intende esaminare il sistema eCall seguendo un duplice percorso: da un lato, approfondendo aspetti normativi e tecnici, dall'altro, offrendo una riflessione critica a partire da un caso concreto verificatosi in Italia, che ha evidenziato le fragilità sistemiche legate all'impiego non ottimale della tecnologia in contesti reali.

1. Il sistema eCall

Il sistema eCall è progettato per attivare automaticamente una chiamata di emergenza al numero unico europeo 112 in caso di collisione grave, trasmettendo un *minimum set of data* ai *Public Safety Answering Points* (PSAP), ossia i centri di risposta per le emergenze. Tali dati comprendono posizione GPS, tipo di veicolo, orario dell'incidente, direzione di marcia e modalità di attivazione della chiamata. Il sistema si compone di tre elementi principali: il modulo IVS (In-Vehicle System), il ricevitore GNSS ed il modem cellulare GSM/UMTS/LTE.

Dal punto di vista normativo, il Regolamento UE 2015/758 ha imposto l'obbligatorietà dell'eCall nei veicoli nuovi dal 2018, fissando requisiti minimi di interoperabilità, disponibilità ed accuratezza. Il sistema è inoltre soggetto al rispetto del Regolamento generale sulla protezione dei dati, poiché richiede il trattamento di informazioni personali sensibili.

Negli ultimi anni, il legislatore europeo ha puntato sull'integrazione tra trasporti e tecnologie emergenti, con particolare attenzione all'intelligenza artificiale. L'entrata in vigore dell'Artificial Intelligence Act (Regolamento UE 2024/1689) rappresenta punto di svolta: i sistemi di IA nei veicoli connessi, se ad "alto rischio", sono oggi sottoposti ad obblighi stringenti in termini di trasparenza algoritmica, supervisione umana e responsabilità condivisa. A ciò si affianca la Direttiva 2024/2853, che ha esteso la disciplina sulla responsabilità da prodotto difettoso, includendo software, firmware e componenti algoritmici. Ne risulta un contesto sempre più complesso, in cui

l'eCall non è più un semplice strumento ausiliario, ma parte di un ecosistema veicolare intelligente e normativamente vigilato. Qui il ruolo umano resta decisivo come custodia critica, affinché l'automazione non degeneri in disumanizzazione e la precisione algoritmica non prevalga sul diritto alla sicurezza e alla comprensibilità delle scelte.

In tale condizione, la configurazione infrastrutturale nazionale è decisiva per l'efficacia del sistema. In Italia, l'accentramento del servizio in un'unica centrale operativa a Varese è un aspetto vulnerabile, per tempestività e conoscenza delle specificità territoriali. In Germania operatori privati filtrano il segnale prima dell'inoltro ai PSAP pubblici, riducendo i falsi allarmi e personalizzando il servizio. In Francia un modello ibrido affida ai costruttori automobilistici dispositivi eCall collegati a centrali private che cooperano costantemente con i PSAP statali, garantendo maggiore resilienza nelle aree rurali e montane.

Il confronto tra i modelli europei mostra differenze significative: in Germania, la presenza di operatori privati ha ridotto i falsi allarmi al 18% contro il 25-40% dei sistemi centralizzati come quello italiano (Rapporto Commissione Europea, 2023); in Francia, il modello ibrido ha comportato una riduzione media dei tempi di intervento del 12% nelle aree rurali e una diminuzione del 20% dei costi legati a falsi allarmi (Ministero dell'Interno francese, 2022).

Questi dati evidenziano come modelli più flessibili e decentrati garantiscano migliori performance, sia nella riduzione dei falsi allarmi sia nella tempestività ed appropriatezza degli interventi.

2. Il Caso Temennotte

Il 5 settembre 2024, nella località montana di Temennotte, frazione del comune di Sant'Agapito (IS), si è verificato un episodio emblematico delle criticità dei sistemi eCall. La centrale di Varese ha ricevuto un segnale di emergenza da un veicolo che, secondo l'algoritmo, aveva subito una collisione ad alta intensità. Attivato il protocollo di urgenza, furono mobilitati il Corpo Nazionale del Soccorso Alpino, un'unità del 118 ed un elicottero di soccorso.

Gli operatori trovarono il veicolo integro e non immatricolato, in sosta in un'area impervia senza copertura stabile. La localizzazione si era rivelata imprecisa ed il sistema IVS, probabilmente a causa di un'anomalia meccanica o di un urto non dannoso, aveva attivato erroneamente l'allarme.

L'episodio, riportato dalla stampa locale (TVI Molise, 2024), evidenzia i limiti di affidabilità dei sistemi informatici in situazioni di emergenza e l'urgenza di migliorarne efficienza e controlli. Una priorità sarebbe la dislocazione della centrale di acquisizione del segnale a livello regionale: la centrale di Varese, priva di conoscenza dettagliata del territorio, rischia di rallentare i soccorsi invece di accelerarli.

L'invio del segnale, classificato come "collisione ad alta intensità", ha attivato il protocollo di massima urgenza, ma una volta sul posto il Soccorso Alpino ha rilevato non solo l'assenza di un incidente, ma anche l'inattendibilità della posizione: il veicolo era in sosta da giorni, in un punto impervio e privo di copertura stabile.

Le conseguenze operative: l'elisoccorso ha comportato costi rilevanti a carico della sanità pubblica regionale, mentre le squadre a terra hanno affrontato rischi in un territorio montano ed in condizioni climatiche avverse. Tali imprecisioni pongono interrogativi giuridici sulle responsabilità civili e penali da malfunzionamenti: la difficoltà di localizzazione incide sui tempi di intervento e sui costi, con rischio di impiego improprio di risorse pubbliche.

Il caso Temennotte non è isolato: studi recenti stimano un tasso di falsi positivi tra il 25% e il 40%, soprattutto in aree geograficamente complesse; la mancanza di un "indice di affidabilità algoritmica" limita la capacità dei PSAP di stabilire priorità d'intervento.

L'introduzione di componenti di apprendimento automatico nei sistemi IVS potrebbe migliorare l'accuratezza, distinguendo tra eventi reali e perturbazioni non rilevanti... Moduli diagnostici a bordo potrebbero trasmettere dati biometrici o ambientali (temperatura, pressione, suoni anomali), integrati da un filtro umano supervisionato, offrendo una seconda soglia di verifica prima della mobilitazione delle risorse di emergenza.

3. Il dibattito sulla natura della responsabilità dell'intelligenza artificiale

La questione della responsabilità derivante dall'uso dell'intelligenza artificiale è al centro dei dibattiti interdisciplinari degli ultimi anni. Sebbene la Direttiva 2024/2853 rappresenti un primo passo normativo, il suo impatto pratico resta incerto e, al di fuori di questa regolamentazione, l'ordinamento legislativo rimane lacunoso.

Uno dei temi principali riguarda l'eventuale attribuzione di capacità legale e personalità giuridica all'IA, ossia se essa possa essere considerata responsabile in modo autonomo (Pretti, 2020). Alcuni propongono

l'equiparazione dei robot alle persone giuridiche, ma la tesi dominante la esclude per le implicazioni etico-giuridiche. In ordinamenti come quello italiano sorgerebbero problematiche irrisolte, ad esempio sull'individuazione di un soggetto tenuto al risarcimento o sul riconoscimento di un danno morale ad un robot. Un'interpretazione alternativa collega l'IA all'istituto della rappresentanza (art. 1387 c.c.), distribuendo la responsabilità tra utilizzatore e contraente senza attribuire capacità giuridica autonoma ai sistemi informatici (Teubner, 2019).

Un caso singolare è quello dell'Arabia Saudita, che nel 2017 ha concesso la "cittadinanza" al robot umanoide Sophia (Cuthbert, 2017); qui, la personalità giuridica è concepita come un insieme modulabile di diritti e doveri, fermo restando che i diritti umani restano inalienabili (Alqodsi, Gura, 2023). Ciò suggerisce che la concezione di personalità giuridica possa evolvere con lo sviluppo dell'IA.

Tuttavia, nel breve e medio termine il riconoscimento di una soggettività giuridica autonoma per l'IA appare improbabile. Le norme attuali si concentrano su sviluppatori ed utilizzatori, escludendo l'attribuzione di diritti o doveri ai sistemi, considerati strumenti tecnologici e non soggetti di diritto.

3.1. La responsabilità civile

In Italia, la responsabilità per prodotto difettoso è disciplinata dal D. Lgs. 206/2005, in attuazione della Direttiva 85/374/CEE. L'art. 114 cod. cons. definisce difettoso un prodotto che non assicura il livello di sicurezza atteso, mentre l'art. 115 attribuisce la responsabilità al produttore, salvo prova contraria. Tale disciplina si applica anche ai sistemi tecnologici installati nei veicoli: la giurisprudenza prevalente conferma la responsabilità primaria del produttore, mentre la responsabilità del fornitore sussiste solo se il produttore sia sconosciuto o non collabori, principio esteso anche ai software di navigazione e localizzazione. La Cassazione ha ribadito che i consumatori hanno diritto alla riparazione o sostituzione gratuita di veicoli con componenti difettose e che il produttore può essere esonerato solo se dimostra l'imprevedibilità del difetto secondo le conoscenze tecnico-scientifiche disponibili. Così, un GPS che malfunzioni in zone montuose potrebbe costituire un difetto se compromette l'adeguatezza tecnica del sistema. In casi come quello di Temennotte, un difetto che ritardi i soccorsi può configurare responsabilità diretta per mancata conformità alle legittime aspettative del consumatore.

La dottrina sottolinea che il difetto va valutato in relazione alle aspettative del consumatore ed agli standard tecnici di riferimento, pertanto, i produttori di sistemi come i GPS devono garantire sicurezza e affidabilità conformi alle attese, rispondendo dei difetti che compromettano il funzionamento (Rumi, 2024).

3.2. La responsabilità penale da malfunzionamento dei sistemi informatici nelle automobili

L'articolo 658 c.c. punisce «chiunque, annunziando disastri, infortuni o pericoli inesistenti, suscita allarme presso l'Autorità, o presso enti o persone che esercitano un pubblico servizio, è punito con l'arresto fino a sei mesi o con l'ammenda da euro 10 a euro 516» (Gazzetta Ufficiale – Codice penale). La norma tutela la tranquillità pubblica contro i falsi allarmi (Caringella, 2016). Si tratta di una contravvenzione che richiede dolo generico, ossia la volontà di annunciare un pericolo inesistente: non è sufficiente un malfunzionamento tecnico di dispositivi come un GPS, che manca dell'elemento soggettivo. Tuttavia, se l'errata segnalazione derivasse da uso negligente o mancato controllo, si potrebbe ipotizzare responsabilità a titolo di colpa.

Il malfunzionamento può essere valutato anche alla luce dell'art. 340 c.p., che punisce l'interruzione o turbamento di un pubblico servizio. La giurisprudenza ammette che il reato possa configurarsi anche in forma colposa, quando l'interruzione deriva da negligenza, imprudenza o imperizia, purché vi sia un nesso causale con l'evento.

In caso di guasto tecnico di un sistema IA che generi un falso allarme e attivi impropriamente i servizi di emergenza, si distinguono due ipotesi: se il malfunzionamento è volontario o manipolato, si configura il reato di interruzione di pubblico servizio; se deriva da negligenza o imperizia (es. omessa manutenzione), può emergere una responsabilità colposa, purché l'evento fosse prevedibile ed evitabile. Se il malfunzionamento è dovuto a un evento imprevedibile e non imputabile, si esclude la responsabilità penale.

In conclusione, la responsabilità penale per malfunzionamento dei sistemi informatici nei veicoli dipende da una valutazione caso per caso, considerando la presenza di dolo o colpa e l'effettivo nesso causale tra il malfunzionamento e l'evento dannoso.

4. Implicazioni etiche nell'uso dell'intelligenza artificiale nei sistemi di sicurezza

L'intelligenza artificiale applicata al settore automotive pone interrogativi etici che non possono essere ridotti alla sola efficienza tecnica o al rispetto normativo. Seguendo la prospettiva di una "algoretica", proposta da Paolo Benanti, occorre interrogarsi sulla capacità degli algoritmi di rispettare la dignità umana e i principi di giustizia, trasparenza e responsabilità (Benanti, 2018). Le decisioni devono riflettere valori condivisi ed essere comprensibili.

L'etica dell'IA non può essere ricondotta ad una mera somma di regole o ad una logica utilitaristica. L'azione morale implica deliberazione e responsabilità, qualità proprie dell'essere umano, mentre i sistemi automatizzati restano strumenti privi di intenzionalità: la loro "autonomia" è funzionale e non morale. Perciò la responsabilità ultima delle scelte algoritmiche deve restare in capo a progettisti, produttori, utenti e istituzioni, secondo un principio di "riserva d'umanità".

La sfida, come sottolineato anche dalle linee guida europee sull'IA affidabile e dall'AI Act, consiste nel garantire la trasparenza delle logiche decisionali, possibilità di controllo umano e rendicontazione sociale delle scelte algoritmiche. È necessario che i sistemi nei veicoli siano progettati "by design" secondo criteri di equità ed inclusività, evitando che bias culturali o tecnici si traducano in discriminazioni.

Il "trolley problem"[1], molto discusso nell'etica applicata, mostra i limiti di una standardizzazione automatica delle scelte morali: i dilemmi posti dalla guida automatizzata non sono risolvibili attraverso calcoli probabilistici, ma richiedono una riflessione pubblica e democratica sui valori da incorporare nella tecnologia. Anche il "Moral Machine Experiment"[2] e le ricerche empiriche condotte[3] dimostrano che le preferenze morali variano tra culture e situazioni, e che la fiducia degli utenti nelle decisioni delle macchine è ancora fragile. Da qui l'importanza di un approccio multilivello, che includa

[1] Un esperimento mentale che presenta un dilemma morale: un tram fuori controllo sta per investire cinque persone legate sui binari, ma è possibile deviarlo su un altro binario dove si trova una sola persona legata. La questione è se sia moralmente lecito intervenire per salvare i cinque sacrificando uno. Questo problema fu introdotto dalla filosofa britannica Philippa Foot nel 1967. Successivamente, la filosofa americana Judith Jarvis Thomson ampliò l'analisi del dilemma nel 1976, contribuendo a diffondere il dibattito accademico sul tema.

[2] https://www.moralmachine.net.

[3] https://docs.google.com/forms/d/1Zpr5Mxms9niNHSlnu3r7ghabg6nLoBn2E1O1tz_kR28/edit?pli=1&authuser=1#responses.

audit indipendenti, consultazione pubblica e formazione continua degli attori coinvolti.

Infine, come suggerito dalla riflessione sull'"algorpolitica", la questione etica riguarda non solo i singoli algoritmi, ma la governance complessiva delle innovazioni tecnologiche e la loro capacità di servire il bene comune. L'obiettivo deve essere un'IA nei veicoli non solo sicura e affidabile, ma anche giusta, comprensibile e socialmente legittimata.

Conclusioni

Le sfide poste dall'implementazione dei sistemi eCall e dall'introduzione dell'IA nella gestione delle emergenze stradali richiedono un ripensamento dell'architettura attuale, alla luce dei dati e delle esperienze maturate in Europa. L'analisi dimostra che modelli più flessibili e decentrati, con filtri iniziali affidati anche ad operatori privati e una maggiore collaborazione pubblico-privato, migliorano l'efficacia del processo, riducendo falsi allarmi e tempi di intervento, con benefici in termini di risorse.

In questa prospettiva, anche in Italia sarebbe auspicabile sperimentare sistemi decentrati, affiancati dall'ottimizzazione delle tecnologie di bordo in grado di fornire dati più affidabili al momento della chiamata di emergenza. La trasparenza e la responsabilità nella gestione dei dati e delle decisioni automatizzate dovrebbero essere garantite da audit regolari e report pubblici, anche per accrescere la fiducia nelle nuove tecnologie.

Formazione degli operatori e linee guida etiche sono indispensabili per una gestione appropriata delle emergenze, nel rispetto dei principi europei di trasparenza e responsabilità. In questa cornice è indispensabile la raccolta sistematica di dati comparativi tra gli Stati membri, così da diffondere le migliori pratiche e favorire una convergenza normativa ed operativa.

Solo un approccio integrato, capace di coniugare innovazione tecnologica, responsabilità umana e trasparenza, potrà garantire sistemi di emergenza efficaci, equi e sostenibili, capaci di tutelare i diritti fondamentali anche nelle situazioni più critiche.

Riferimenti bibliografici

Al Mureden E. (2024). *Diritto dell'automotive*. Bologna: il Mulino.

Awad E., Dsouza S., Kim R. et al. (2018). The Moral Machine experiment. *Nature*, 563: 59-64. https://doi.org/10.1038/s41586-018-0637-6.

Benanti P. (2018). *Oracoli. Tra algoretica e algocrazia*. Roma.

Benanti P. (2024). Ecco cos'è l'algoretica e perché ce n'è bisogno. Milano. https://adeccogroup.it/paolo-benanti-algoretica-cosa-e/.

Benedetti F. (2020). La responsabilità civile nei sistemi di guida autonoma: una sfida per il diritto contemporaneo. *Rivista di Diritto Privato*, 4: 321-345.

Brocardi. https://www.brocardi.it/codice-penale/libro-secondo/titolo-ii/capo-ii/art340.html.

Calabresi G., Al Mureden E. (2020). Driverless car e responsabilità civile. *Rivista di Diritto Bancario*, supplemento gennaio/marzo. https://rivista.dirittobancario.it/driverless-car-e-responsabilita-civile (visitato il 22 giugno 2025).

Calabresi G., Al Mureden E. (2021). *Driverless cars*. Bologna: il Mulino.

Caringella F. (2016). *Manuale di diritto penale. Parte speciale*. Napoli: Dike Giuridica.

Cenci D. (2020). Responsabilità da prodotto e sistemi autonomi: nuove frontiere della responsabilità civile. *Il Diritto dell'Informazione e dell'Informatica*, 4(2): 305-332.

Cruciani A. (2022). Hacker e auto, violare una vettura è fin troppo facile. I rischi e come difendersi. *Corriere della Sera*, 19 giugno. https://www.corriere.it/tecnologia/22_giugno_19/hacker-auto-violare-vettura-auto-fin-troppo-facile-rischi-come-difendersi-91adca52-efa7-11ec-8f59-93717c23f0aa.shtml (visitato il 23 giugno 2025).

Cuthbert O. (2017). Saudi Arabia becomes first country to grant citizenship to a robot. *Arab News*. https://www.arabnews.com/node/1183166/saudi-arabia.

Della Giustina C., De Gioia Carabellese P. (2023). Il futuro ruolo dell'assicuratore nei rischi legali dei veicoli automatici: unmanned vehicles, trolley problems and data protection. *Rivista Trimestrale di Diritto e Procedura Civile*, 4: 1235. Milano.

European Commission (2023). *Impact Assessment Report on the Artificial Intelligence Act*. Bruxelles: Commissione Europea. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13140-Artificial-Intelligence-Act.

Faralli C. (2019). Diritti e nuove tecnologie. In *Rivista di Scienze della Comunicazione*.

Ferrari V. (2020). Note socio-giuridiche introduttive per una discussione su diritto, intelligenza artificiale e big data. *Sociologia del Diritto*, 3: 9-32. DOI: 10.3280/SD2020-003001.

Ferrari V. (2021). Nuove tecnologie – diritto – impatto reciproco – potenzialità – discrasie. In *Diritto e nuove tecnologie della comunicazione*, Adunanza del 6 maggio 2021.

Ferrari V. (2022). Diritto robotizzato? Riflessioni socio-giuridiche sulle nuove tecnologie della comunicazione. *Annali del Dipartimento Giuridico dell'Università degli Studi del Molise*: 3-14.

Ferrari V. (2024). Experiences and queries about artificial intelligence and law. Oral contribution to a panel on law and artificial intelligence at the annual meeting of the ISA Research Committee in Sociology of Law, Bangor University, 3-7 settembre 2024.

Galletti M., Zipoli Caiani S. (2024). *Filosofia dell'intelligenza artificiale*. Bologna.

Gill T. (2021). Ethical dilemmas are really important to potential adopters of autonomous vehicles. *Ethics and Information Technology*.

Miniscalco N. (2024). *L'intelligenza artificiale in movimento. Impatto sui diritti costituzionali*. Milano.

Montani V. (2024). Intervista al Prof. Luca Grion. Etica delle macchine e responsabilità umana: la morale al tempo dell'intelligenza artificiale. *Diritto Mercato e Tecnologia*. https://www.dimt.it/news/intervista-al-prof-luca-grion-etica-delle-macchine-e-responsabilita-umana-la-morale-al-tempo-dellintelligenza-artificiale/.

Pezzano G. (2022). *L'etica delle macchine spiegata in 100 minuti (attraverso i veicoli autonomi)*. Torino.

Pretti G. (2020). *La responsabilità nell'intelligenza artificiale e nella robotica*. Milano.

Rubin A., Bonazzi M., Mancini M., Mattioli G. (a cura di) (2024). *Veicoli a guida autonoma. Opportunità, sfide e prospettive future di una tecnologia per una mobilità sicura, efficiente e sostenibile*. Udine: Mimesis Edizioni.

Rumi T. (2024). *La product liability nell'era dell'I.A.* Napoli.

Sartor G. (2022). *L'intelligenza artificiale e il diritto*. Torino: Giappichelli.

Scagliarini S. (2023). Smart roads e driverless cars: tra diritto, tecnologie, etica pubblica. In *Diritto e vulnerabilità – Studi e ricerche del CRID*. Torino.

Simmel G. (1968). *L'etica e i problemi della cultura moderna*. Napoli: Alfredo Guida Editore, 2004 (trad. it.).

Society of Automotive Engineers (2021). Livelli di automazione. https://www.sae.org/news/2021/06/sae-revises-levels-of-automation (visitato il 23 giugno 2025).

Schartel S. (2023). *Artificial intelligence and the ethics of self-driving cars: Societal implications*. Cambridge: MIT Press.

Teubner G. (2019). Soggetti giuridici digitali? Sullo status privatistico degli agenti software autonomi. A cura di Pasquale Femia. Napoli: Università degli Studi della Campania "Luigi Vanvitelli", Dipartimento di Scienze Politiche Jean Monnet.

TVI Molise (2024). GPS impazzito mette in moto la macchina dei soccorsi ma è un falso allarme. https://www.tvimolise.it/gps-impazzito-mette-in-moto-la-macchina-sei-soccorsi-ma-e-un-falso-allarme/.

Teti A. (2025). *Digital profiling. L'analisi dell'individuo tra metodologie, tecniche e intelligenza artificiale*. Milano: Il Sole 24 Ore.

Impatto algoritmico dell'IA nella guida stradale autonoma

di *Antonella Tennenini*[*]

La guida automatica implica un "dialogo" tra le infrastrutture e i veicoli, nonché una connessione con gli utenti vulnerabili della strada. Essa ridurrà l'incidentalità o congestionerà ancora di più il flusso del traffico? Il presente contributo enuclea benefici e criticità degli algoritmi nel settore automobilistico e infrastrutturale, con qualche esempio, come la manutenzione della *Expressway* in Cina, avvenuta mediante gli agenti artificiali.

Parole chiave: sicurezza stradale; guida autonoma; cybersecurity; vulnerabile road users; Smart Road; privacy.

Algorithmic impact of AI in autonomous road driving

Autonomous driving involves a "dialogue" between infrastructure and vehicles, as well as a connection with vulnerable road users. Will it reduce accidents or will it congest the flow of traffic even more? This paper examines the benefits and criticalities of algorithms in the automotive and infrastructure sectors, with some examples, such as the maintenance of the Expressway in China, which took place using artificial agents.

Keywords: road safety; autonomous driving; cybersecurity; vulnerable road users; Smart Road; privacy.

Introduzione

Il presente contributo esamina la guida completamente autonoma su strada e svincolata dall'umano in prospettiva futura, in quanto, allo stato attuale, si è ad un livello più esplorativo che pratico, almeno in ambito europeo. L'approccio teorico non sarà avveniristico, entusiasta o pessimistico, ma ispirato ad uno sguardo critico, volto a comprendere come la costruzione e l'uso delle automobili a base algoritmica possano incidere sugli stili di vita, sugli aspetti etici, economici e giuridici, divenendo un fenomeno sociale innovativo dai risvolti problematici.

DOI: 10.5281/zenodo.17524623

[*] Università degli Studi di Roma Tor Vergata. a.tennenini@alice.it.

Partiamo da qualche dato italiano sul contesto di transizione: ogni anno sulle strade si registrano incidenti talmente impattanti da spazzare via l'equivalente di intere cittadine. Solo nel 2023 le vittime sono state 3.039[1], e nel primo semestre del 2024, 1.429, con un aumento del 4%[2]. Nonostante le trasversali iniziative di prevenzione e le recenti restrizioni del Codice della Strada, la strage di pedoni, ciclisti, monopattinisti e automobilisti, è continua. Fattori endogeni ed esogeni, come la distrazione umana, la velocità troppo elevata, il mancato rispetto della precedenza, e, recentemente, sempre più l'uso dello *smarthphone*, la "sbornia del terzo millennio", il consumo di alcol e di sostanze stupefacenti, rappresentano le preminenti cause dei sinistri. La prevenzione non può passare solo per forme repressive e punitive, a colpi di draconiane sanzioni, che pure sono necessarie per l'effetto deterrente, ma è soprattutto l'insieme dei processi educativi e formativi, sin dall'età scolare infantile, prima dell'acquisizione della patente, che possono contribuire alla maturazione di comportamenti virtuosi, tali da indurre i cittadini a fare scelte più consapevoli e caute nel frequentare la strada. Puntare sull'educazione stradale significa aumentare la qualità della sicurezza e ridurre il più possibile i pericoli per la salute, con particolare riguardo ai pedoni, utenti "vulnerabili" della strada, che, in caso di incidente, sono soggetti ad un elevato rischio di esito fatale.

Ora, intorno alla mobilità si snoda l'assetto culturale, sociale ed economico del tessuto urbano ed extraurbano, ma soprattutto della quotidianità esistenziale delle persone (Castells, 2003). Nel Novecento la graduale propagazione di massa delle automobili ha ampliato lo svolgimento delle attività umane e soddisfatto molteplici esigenze di lavoro e svago, accorciando le distanze. Dal punto di vista sociale, questo ha comportato trasformazioni antropologiche e cognitive importanti, ad esempio in riferimento alla percezione e all'immersione nello spazio e nel tempo: nella percorrenza di tragitti predefiniti e standardizzati, il conducente deve attenersi a regole uniformi, in contrapposizione ad un'esplorazione sensoriale dei luoghi fatta a piedi o in bicicletta a ritmi più lenti. Quanto enucleato fa riferimento ad un sistema ampiamente intriso di tecnica, complessivamente analogico e tradizionale, sebbene con molti aspetti, almeno dagli anni Duemila, viranti verso gestioni ibride, con l'introduzione del digitale. Oggi il settore della mobilità è investito dalla pervasività dell'IA, il che richiede un'estesa comprensione del nuovo. Le modifiche interessano le relazioni tra macchine a guida *in toto*

[1] https://www.istat.it/wp-content/uploads/2024/07/infografica-incidenti-stradali-2024.pdf
[2] https://www.ilsole24ore.com/art/istat-2024-aumentano-vittime-strada-4percento-auto-record-italia-694-ogni-mille-abitanti-ue-sono-571-AGWF2GgB

automatica (e/o automatizzata), infrastrutture e *vulnerabile road users*, i flussi di traffico con i sistemi di videosorveglianza, la gestione della *cyber-security*, nonché le abitudini e i costumi consolidati. L'uso dell'IA, "capace" di apprendere e agire autonomamente, mediante algoritmi addestrati per mano e menti umane, potrebbe risolvere, migliorare, o, invece, eventualmente peggiorare la compagine infrastrutturale e il sistema della mobilità, così come si sono configurati nel tempo e funzionano tuttora? Ci saranno effetti positivi e negativi dell'impatto algoritmico in un settore di punta dell'economia mondiale, come quello automobilistico, ma anche da un punto di vista etico.

Di seguito si tracceranno argomentazioni sulle indubbie opportunità che l'IA promette, con la dimostrazione di qualche esempio fattivo, ma anche paradossalmente sulle limitazioni che l'umano dovrà fronteggiare, per non lasciarsi sopraffare dalle mirabolanti imprese artificiali.

1. Vantaggi e criticità dell'IA nella mobilità

Gli investimenti dei *brand* automobilistici sull'automazione di molte funzionalità delle macchine sono enormi, fino all'obiettivo della guida autonoma, che, tuttavia, incontra i suoi limiti per problematiche molto più complesse di quelle su cui gli ingegneri stanno attualmente lavorando per l'introduzione degli algoritmi su strada, almeno nel contesto europeo.

I sistemi di guida autonoma, infatti, funzionano agevolmente in determinate aree geolocalizzate delle dimensioni di un quartiere, e in futuro di una città. Le sfide tecniche e le implementazioni economiche necessarie per uscire da queste zone delimitate sono immense. Confidare di risolvere i problemi connessi alla guida con l'IA al momento è preoccupante, poiché il controllo dei veicoli si basa sugli stessi principi di ChatGPT. Questa utilizza il ragionamento statistico (con ampi margini di continuo miglioramento), anziché comprendere la situazione, il contesto o qualsiasi altro fattore che la contingenza umana terrebbe invece in considerazione, pur sempre con l'ampia possibilità di commettere errori, talvolta, però, rimediabili nell'attimo del pericolo imminente per intuizioni spontanee e lontane da ogni sovrastruttura[3]. La capacità decisionale umana, infatti, di vigilare situazioni estreme, che richiedono immediatezza di azione, non può essere paragonata e sostituita dall'uso predittivo di un sistema algoritmico.

[3] https://www.lestradedellinformazione.it/rubriche/le-strade-della-tecnica/guida-autonoma-sviluppo-forte-ritardo

Insomma la guida automatica, progettata per il trasporto di persone e di merci, può raggiungere interessi generali come la sicurezza stradale, la riduzione dell'inquinamento e l'inclusione sociale; infatti, potrebbe fornire supporto, ad esempio, agli anziani e alle persone con situazioni di disabilità, le quali avrebbero difficoltà nell'approcciare i mezzi tradizionali, favorendo così i principi di non discriminazione e pari opportunità di movimento e di socializzazione. Inoltre, i conducenti che avrebbero lunghe percorrenze quotidiane, sarebbero sgravati dall'incombenza della guida e potrebbero fruire dell'abitacolo per riposo, lavoro e svago (Calabresi, 2021).

Questo processo innovativo, dalle grandi potenzialità, è, tuttavia, disseminato di insidie e rischi, per cui sarà necessario garantire la previsione e la prevenzione di errori e di attacchi *cyber*. Il funzionamento automatico ruota intorno all'uso massiccio dei dati, che, se eseguito in modo improprio, può arrecare danni, in particolare se non ne viene assicurata la piena tutela per la vita umana, in un campo in cui le automobili saranno in comunicazione tra di loro, con gli utenti e con l'infrastruttura stradale. Gli infiniti dati e le informazioni consentono al costruttore dell'automobile e ad altri soggetti connessi di mappare l'utilizzatore, codificando con precisione i suoi bisogni, le sue inclinazioni, i suoi gusti, mediante i suoi spostamenti, così da adattare le caratteristiche del veicolo alle sue esigenze e stabilire un rapporto di fidelizzazione (Scagliarini, 2019). Tuttavia, questa modalità ribalta totalmente la facoltà umana di scegliere/modificare e latamente l'interazione tra uomo e macchina: è l'algoritmo che suggerisce all'umano e lo insegue, e non il contrario. A tal proposito, il problema della manomissione, alterazione/difetto o manipolazione del software della *self-driving car* ha effetti sulle scelte etiche e le responsabilità. Queste non farebbero più capo al conducente (com'è tuttora, anche da normativa), ma da un lato ai produttori, e dall'altro agli organi legislativi che permettono la costruzione e l'uso di determinati software e hardware, che su strada possono avere sensori per individuare soggetti e situazioni con alcuni problemi, ma non riescono, di certo, ad esempio, ad identificare i soggetti con fragilità psico-motorie, che magari abbisognano di più tempo e soste per un attraversamento pedonale.

Insomma, la persona che sale su un'automobile automatica ha il diritto ad avere alcune garanzie di sicurezza e di stabilità, così come i pedoni e i ciclisti. Se un conducente umano nella guida tradizionale riesce a sterzare all'improvviso in caso di mancato rispetto di una precedenza ad un incrocio da parte di un altro utente, nei sistemi automatici preimpostati con l'IA, questo sarà forse possibile solo dopo un alto grado di addestramento algoritmico, ma porrà in ogni caso il problema della responsabilità civile e penale in caso di sinistro o lesioni.

Essere alla guida da parte del *sapiens*, infatti, significa combinare e coordinare una serie di operazioni legate ad un'adeguata conoscenza di tutti i dispositivi interni al veicolo e delle norme di comportamento nella circolazione, con l'attenzione e la concentrazione continue al possibile rischio, dato dalle condizioni variabili del contesto viario. Gli agenti artificiali riescono a guidare indubbiamente con più precisione, peraltro scevri da fattori emotivi, ma non hanno volontà e, soprattutto, coscienza del senso dell'imprevedibile e dell'incertezza.

Il Nobel 2024 per la Fisica è stato assegnato a Hopfield e Hinton, pionieri degli studi sulle reti neurali artificiali e sui computer capaci di imparare autonomamente. Il professore canadese Hinton prevede che l'IA

> avrà conseguenze sull'umanità paragonabili alla rivoluzione industriale. Allora le macchine ci superavano in forza fisica, ora sono destinate a superarci dal punto di vista intellettuale. Ci saranno effetti estremamente positivi, avremo una medicina migliore e potremo lavorare con un assistente artificiale che ci renderà più produttivi. Ci potranno essere però anche delle conseguenze negative, qualora le macchine riescano a sfuggire al nostro controllo[4].

Wadhwa (2017), d'altronde, pone questioni similari riguardo all'applicazione dell'IA anche nei trasporti: a chi può andare a beneficio, se promuove l'autonomia o la dipendenza dai mezzi, e se comporta la perdita di tipologie di lavoro, soppiantate da nuove figure professionali; in più, solleva il problema della completa perdita della *privacy*, intesa come protezione della riservatezza esclusiva dell'individuo, soprattutto per i profili più delicati, in riferimento a persone minori di età o che hanno particolari esigenze di salute; e ancora, evidenzia la maggiore disuguaglianza economica, sociale e culturale, immaginando, quindi, un "futuro alienante e spaventoso", ma anche "eugenetico". Va da sé che il cambio di paradigma nella guida non sarà privo di costi e controversie e dovrà fare i conti sia con l'eventuale inadeguatezza cognitiva e sociale, sia con le possibili ritrosie culturali e con la mancanza di fiducia da parte di alcune fasce della popolazione[5].

[4] https://www.repubblica.it/italia/2024/10/08/news/premio_nobel_fisica_2024_john_hopfield_geoffrey_hinton-423542854/?ref=RHLF-BG-P6-S1-T1

[5] Cfr. studi a proposito del senso di fiducia verso la guida automatica in https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2023.1279271/full

Dunque, l'opinione pubblica potrebbe dividersi tra chi riconosce i benefici della guida automatica[6], e chi, invece, non vuole vivere nella cosiddetta "sorveglianza del capitalismo".

Al momento le nostre nuove vetture hanno già diverse funzionalità automatizzate, integrate nei sistemi di *driver assistance* (ADAS): il *cruise control* adattivo, la frenata automatica di emergenza, e il monitoraggio degli angoli ciechi; tuttavia, la transizione verso la completa automazione necessita di un arco temporale più esteso, la cui durata a priori non si può stabilire, sia per consentire le indispensabili modifiche tecniche, come la combinazione di sensori, telecamere, radar, connessi al sistema infrastrutturale e individuale, sia per cercare di attenuare divari di ordine economico e sociale[7]. Le strade, infatti, sono progettate e realizzate per essere percorse dai mezzi tradizionali, per cui vi sarà una fase di adeguamento infrastrutturale, da conciliare con le esigenze ambientali, di certo non secondarie, e poi di condivisione delle stesse infrastrutture tra le macchine tradizionali e quelle a guida automatica e connesse, con conseguenze da non sottovalutare, ancora una volta, sul piano della sicurezza ed etico. Questo rinnovamento richiede elevati costi per essere predisposto, che ad ora sono sostenuti solo da gruppi oligopolistici, molto interessati al solo profitto. Pertanto, vi saranno disuguaglianze sociali, poiché inizialmente le macchine automatiche saranno un privilegio oligarchico, prima eventualmente di diventare accessibili alla moltitudine. Alcune sperimentazioni delle modifiche alla trasportistica riguardano, con gradualità, metropolitane – a Roma la Linea C utilizza il sistema di automazione integrale che sostituisce l'operatività del macchinista alla guida –, taxi – attivi in grandi città, come San Francisco e Wuhan –, *shuttle bus* e grandi camion che fanno da precursori su percorsi stradali dedicati. In ambito privato sarà più difficile introdurre la *driverless car*, poiché le automobili tradizionali, viaggianti su strade talvolta molto bisognose di manutenzione, soddisfano ancora molte esigenze, oltre a simboleggiare la libertà di movimento, l'indipendenza e l'autonomia; esse fanno parte a tutti gli effetti di un costume sociale, di un bisogno esistenziale, ma anche di mode. A tal proposito, uno dei poli di eccellenza per l'innovazione è la "Motor Valley" a Modena, dove sono presenti alcuni tra i più prestigiosi marchi mondiali, che hanno conseguito alti livelli di automazione. Pur tuttavia, la Ferrari si rifiuta di passare alla guida autonoma, al fine di preservare le straordinarie

[6] Cfr. a tal riguardo studi empirici sul panorama internazionale in https://deepblue.lib.umich.edu/handle/2027.42/108384

[7] https://aci.gov.it/onda-verde/n-55-settembre-ottobre-2024/

emozioni date dalle prestazioni vissute al volante dei suoi veicoli, da sempre simbolo di brio e libertà[8].

Analizzate le potenzialità e alcune criticità della guida automatica, passiamo al quadro normativo ed attuativo in Italia.

Nel 2018 il Ministero Infrastrutture e Trasporti ha emanato il decreto cosiddetto *Smart Road*[9], che dispone l'ammodernamento e l'adeguamento tecnologico di tutta la rete stradale attraverso la digitalizzazione delle infrastrutture stradali, anche a supporto di veicoli connessi e con più avanzati livelli di assistenza automatica alla guida. Va evidenziato che l'allineamento fattivo di quanto normato necessita ancora di tempo per concretizzarsi ed espandersi. Più avanti sarà riportato un caso.

Oltre all'indispensabile sistemazione computazionale delle infrastrutture, tuttavia, la normativa dovrà aggiornarsi su molte questioni aperte relative alla guida automatica, con nuovi diritti specifici: su tutte, come già evidenziato, le responsabilità civile e penale in caso di trasgressione delle norme di circolazione. Attualmente, il Codice della Strada non è stato ancora rivisto sulla peculiarità della materia, pertanto il punto di riferimento sono le norme primarie, tra cui l'art. 2054 del Codice Civile che chiama in causa il conducente umano e il proprietario del veicolo in caso di sinistri stradali. Tuttavia, la giurisprudenza dovrà occuparsi, altresì, delle questioni assicurative, dei rischi di "hackeraggio" e di *privacy*. Quest'ultima, già regolata in ambito europeo dal GDPR 2016, rappresenta un vero dilemma, data la condivisione e connessione continue e costanti di dati individuali da tutelare, usati dagli algoritmi su strada.

2. Qualche recente esempio applicativo dell'IA su strada

Nell'agosto del 2010 la *National Expressway*, la strada che in Cina collega Pechino, a nord del Paese, con Hong Kong e Macao, a sud, era diventata famosa per un ingorgo stradale con una coda di circa 100 chilometri per nove giorni. Si tratta di un tratto di 158 chilometri che è stato interessato di recente da importanti lavori di manutenzione: è stato completamente riasfaltato con l'ausilio di droni e macchine robot senza l'intervento di operai, ma solo con la supervisione umana durante le attività e le fasi lavorative. Il lavoro sulla *Expressway* è stato realizzato con un "pavimentatore automatico a guida autonoma" largo 20 metri, che è stato puntato già da molti investitori stranieri

[8] https://www.repubblica.it/tecnologia/2022/06/19/news/niente_guida_autonoma_siamo_la_ferrari-354310675/

[9] Vd. D.M. del 28.02.2018 in G.U. Serie Generale n. 90 del 18.04.2018.

del settore ingegneristico. Infatti, il macchinario è stato in grado di stendere un unico nastro di asfalto per tutta la sua larghezza, migliorando la levigatezza e l'integrità strutturale della strada. Sul tratto autostradale sono stati attivati sei rulli compressori, da 13 tonnellate e sempre senza conducente, per stirare l'asfalto appena posato e tre rulli da 30 tonnellate l'uno. Il loro movimento seguiva uno schema preimpostato dagli algoritmi dell'IA e da una rete di rilevazioni satellitari per misurare con la tolleranza di un centimetro lo spessore dell'asfalto, evitando così di avere angoli di catrame poco compressi e garantendo la qualità e l'uniformità del manto stradale. Inoltre, nello stesso tempo alcuni droni, con lenti ottiche e sensori ad altissima definizione, hanno controllato la presenza di cartellonistica e di tratti più o meno ammalorati, che necessitavano di manutenzione e nuovo asfalto. Questa attività dell'IA ha ridotto notevolmente gli infortuni sul lavoro, in quanto tutte le macchine erano dotate dei cosiddetti "cancelli digitali", ossia sensori che le facevano spegnere in caso di ostacolo, umano e non, che si frapponeva sulla loro strada e nei loro radar a meno di un metro. Tale processo ha coinvolto un *team*, dai 5 ai 20 lavoratori, che hanno sorvegliato da remoto il regolare svolgimento dei lavori, quindi con una forte riduzione del fabbisogno di manodopera. Inevitabile qui domandarsi chi perde lavoro in questi settori come e dove sarà re-impiegato[10].

Passiamo ad un caso italiano. L'Anas sta sperimentando un sistema di *Smart Road* in uno dei tratti autostradali più difficoltosi, la A2 – Autostrada del Mediterraneo, dove ha completato la "*Green Island*" nell'area di parcheggio Contessa Soprana tra gli svincoli di Montalto e Torano, in provincia di Cosenza. Allo stato attuale, sono stati completati gli interventi su circa 140 chilometri lungo i tratti autostradali tra Campania, Basilicata e Calabria, con l'installazione già di 800 postazioni polifunzionali, dotate di telecamere, sistema di IA e sensori che inviano 24 ore su 24 dati alla sala di controllo, collegati tra loro da fibra ottica. Sono in corso i lavori di realizzazione di ulteriori 94,5 chilometri. Il progetto permette ai veicoli in transito di ricevere in tempo reale le informazioni di servizio e/o sicurezza, riguardanti gli incidenti, le code, i rallentamenti, i cantieri, che ad ora sono veicolate tramite media, Isoradio CCISS e pannelli a messaggio variabile[11].

[10]https://www.lestradedellinformazione.it/rubriche/le-strade-della-tecnica/lintelligenza-artificiale-utilizzata-lavori-stradali-cina

[11] https://www.stradeanas.it/it/a2-"autostrada-del-mediterraneo"-completata-la-green-island-della-smart-road-montalto-uffugo

Conclusioni

La guida automatica si sta realizzando, amplificando i possibili vantaggi, ma anche la complessità sociale e le complicazioni gestionali, etiche e finanziarie ad essa correlate. Le *performance* delle automobili automatiche saranno sempre più simili a quelle umane, incluse in futuro le loro abilità e facoltà, previsionali e sensitive. Per questo, si pone il problema della centralità dell'umano (Bertolaso, 2023), del suo controllo sulla tecnica, e non il contrario, all'interno di un disegno più ampio, dove le interazioni e le relazioni saranno sempre più artificiali. Eticisti, ingegneri, giuristi, sociologi devono collaborare per la formulazione di una nuova etica nella, della e per la gestione algoritmica che, rifacendosi ai fondamenti classici, contemperi l'agire umano e artificiale in una coesistenza che necessita di nuovi equilibri, usi e costumi, al di là della regolamentazione, de-regolamentazione, e dei codici deontologici. La nuova etica dovrebbe fondarsi sui principi di trasparenza, di leale concorrenza e di redistribuzione, e mirare alla chiara declinazione di responsabilità (e di conseguenze) nei produttori e consumatori per dirimere inevitabili disorientamenti.

Questo con particolare riguardo alla guida autonoma, che sarà sempre più dialogante e interdipendente tra umano e non umano, connessa ed elettrificata, il che impone di preoccuparsi anche di aspetti ecologici: del consumo energetico, di acqua necessaria per il raffreddamento dei sistemi, di dismissione di tutti i mezzi prodotti finora in circolazione o fermi. Peraltro valutare se è davvero necessaria una sostituzione di tutto ed a tutti i costi e per quali finalità. L'umano è responsabile delle azioni sull'ambiente, sin troppo sfruttato e vessato.

Occorrono educazione, formazione, e partecipazione attiva da parte dei cittadini per il corretto funzionamento automobilistico ed infrastrutturale tramite l'IA (Bergonzini, 2024), al fine di condividere la stessa strada, e magari con mezzi ben diversi.

Riferimenti bibliografici

Bergonzini G. (2024). Sicurezza della città, tecnologie digitali e intelligenza artificiale: tra regole europee, garanzie costituzionali e autonomia locale. *Federalismi.it – Rivista di diritto pubblico italiano, comparato, europeo*, 25: 1-43.

Bertolaso M. (2023). *Umanesimo tecnologico. Una riflessione filosofica sull'intelligenza artificiale*. Roma: Carocci

Calabresi G., Al Mureden E. (2021). *Driverless cars: intelligenza artificiale e futuro della mobilità*. Bologna: Il mulino.

Castells M. (2003). *La città delle reti*. A cura di M. Panarari, C. Rizzo. Milano: Reset.

Scagliarini S. (2019). *Smart roads e driverless cars: tra diritto, tecnologie, etica pubblica*. Torino: Giappichelli.

Wadhwa V. (2017). *Il pilota nell'auto senza pilota: come non perdere il controllo delle nostre vite nell'era dell'intelligenza artificiale*. Milano: LSWR.

Sitografia (Ultima consultazione siti: 24.02.2025)

https://deepblue.lib.umich.edu/handle/2027.42/108384

https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2023.1279271/full. DOI: 10.3389/fpsyg.2023.1279271

https://www.istat.it/wp-content/uploads/2024/07/infografica-incidenti-stradali-2024.pdf

https://www.ilsole24ore.com/art/istat-2024-aumentano-vittime-strada-4percento-auto-record-italia-694-ogni-mille-abitanti-ue-sono-571-AGWF2GgB

https://www.lestradedellinformazione.it/rubriche/le-strade-della-tecnica/guida-autonoma-sviluppo-forte-ritardo

https://www.repubblica.it/italia/2024/10/08/news/premio_nobel_fisica_2024_john_hopfield_geoffrey_hinton-423542854/?ref=RHLF-BG-P6-S1-T1

https://aci.gov.it/onda-verde/n-55-settembre-ottobre-2024/

https://www.repubblica.it/tecnologia/2022/06/19/news/niente_guida_autonoma_siamo_la_ferrari-354310675/

https://www.lestradedellinformazione.it/rubriche/le-strade-della-tecnica/lintelligenza-artificiale-utilizzata-lavori-stradali-cina

https://www.stradeanas.it/it/a2-"autostrada-del-mediterraneo"-completata-la-green-island-della-smart-road-montalto-uffugo

Cmd/Ctrl. Considerazioni etiche e operative sull'uso di sistemi di intelligenza artificiale in supporto alle decisioni militari

di *Michele Carlo Tripeni* [*]

L'uso crescente di sistemi di supporto decisionale basati sull'intelligenza artificiale nelle operazioni militari promette maggiore efficienza e rapidità, ma introduce il rischio dell'*automation bias*. Questo fenomeno può compromettere il giudizio umano e il rispetto del diritto internazionale umanitario. Il saggio esamina tali sfide e propone strategie per bilanciare l'uso dell'IA con il mantenimento del controllo umano.

Parole chiave: AI-DSS; automation bias; etica militare; processo decisionale; controllo umano; diritto internazionale umanitario.

Cmd/Ctrl. Ethical and operational considerations on the use of artificial intelligence systems to support military decisionmaking

The increasing use of Artificial Intelligence-based Decision Support Systems in military operations promises enhanced efficiency and speed but raises concerns about automation bias. This phenomenon may undermine human judgment and compliance with International Humanitarian Law. This paper explores these challenges and proposes strategies to balance AI advantages with the essential role of human oversight in military decision-making.

Keywords: AI-DSS; automation bias; military ethics; decision-making; meaningful human control; laws of war.

Introduzione

Con il crescente utilizzo nelle operazioni militari di sistemi di supporto decisionale basati sull'intelligenza artificiale (AI-DSS), vengono promessi miglioramenti senza precedenti in termini di efficienza, rapidità ed efficacia (tra gli altri: Cook, 2021; Harper, 2018). In effetti, questi sistemi sono in grado di setacciare enormi quantità di informazioni e di fornire raccomandazioni in tempo reale. Funzioni che attualmente sono essenziali per alleviare

DOI: 10.5281/zenodo.17524653

[*] University of Glasgow. michelec.tripeni@gmail.com.

il carico cognitivo del personale militare, il quale opera in contesti sempre più complessi e interconnessi. Tuttavia, questo utilizzo dell'intelligenza artificiale (IA) comporta anche notevoli sfide. Di particolare importanza nel contesto degli AI-DSS sono le problematiche legate all' *automation bias*. Ovvero, il fenomeno per cui gli operatori umani preferiscono deferire alle raccomandazioni generate dai sistemi automatizzati senza verificarle né interrogarle in modo critico. Tale dipendenza dall'IA nel contesto militare può introdurre vulnerabilità operative ed etiche di grande impatto, particolarmente per le loro conseguenze sul rispetto del diritto internazionale umanitario (DIU).

Una delle preoccupazioni più urgenti è proprio che il fenomeno, offuscando il giudizio etico umano degli operatori, possa portare a violazioni dei principi fondamentali del diritto bellico. Questi principi richiedono la valutazione di fattori morali ed operativi complessi da parte dei comandanti. Valutazione che anche i più avanzati sistemi di IA non sono in grado di compiere. In uno scenario di eccessiva dipendenza dalle raccomandazioni dell'IA, dunque, il rispetto delle norme di DIU e il controllo da parte dei comandanti militari potrebbero risultare sostanzialmente erosi. Portando così ad una situazione in cui gli operatori umani, e non tanto i "killer robot" al centro di molti dibattiti internazionali, vengono ridotti a fredde macchine assassine (Carpenter, 2024).

Questo saggio sostiene che il giudizio e la responsabilità finale devono risiedere negli operatori umani come garanzia indispensabile nelle applicazioni militari degli AI-DSS. Propone inoltre, sulla base della letteratura in materia, un approccio equilibrato che concili i vantaggi dell'IA in termini di efficienza ed efficacia con il mantenimento del controllo umano, fattore centrale per il comando militare e per la condotta legale della guerra. Il saggio fornisce inizialmente un breve resoconto dello sviluppo e dell'impiego di AI-DSS in ambito militare. Successivamente analizza la letteratura riguardo i fenomeni cognitivi legati al loro impiego, in particolare l'*automation bias*, e il loro potenziale impatto sul rispetto del DIU. Si conclude poi con alcuni suggerimenti per mitigare questo impatto e proteggere l'autonomia decisionale umana, oltre a fornire ulteriori spunti di riflessione per future ricerche nel campo.

1. Contesto e sviluppo degli AI-DSS nelle operazioni militari

I sistemi informatizzati per il supporto alle decisioni (*decision support systems*, DSS) non sono certamente una novità, né lo è l'idea di impiegare

tecniche di IA a questo scopo. Infatti, a partire almeno dalla metà degli anni Ottanta, nell'ambito del management e dell'informatica si è sviluppato il filone di ricerca sugli *intelligent decision support* systems (termine la cui genesi è attribuita a Holsapple e Whinston (1987)). Questi sistemi verranno successivamente ribattezzati *AI decision support systems* (AI-DSS) con l'affermarsi dell'intelligenza artificiale nel linguaggio comune. Già in questo contesto vi era la consapevolezza che per progettare al meglio tali sistemi fosse necessario comprendere gli utenti e le loro "limitazioni", oltre ai meccanismi secondo cui essi stessi determinassero quale metodo utilizzare per raggiungere una decisione. In questo senso l'obiettivo doveva essere presentare le informazioni in modo congruente con i processi decisionali degli utenti, aiutando semplicemente a scegliere lo strumento più adeguato e, al contempo, ad evitare errori nella manipolazione ed interpretazione dei dati. (Remus, Kottemann, 1986)

In ambito militare lo sviluppo degli AI-DSS ha una lunga storia parallela a quella delle applicazioni civili. I primi sistemi si focalizzavano su applicazioni di logistica nonché su automazioni di tipo *rule-based*, per poi includere sistemi IA sempre più avanzati andando a ricoprire funzioni di analisi d'intelligence e individuazione degli obiettivi. L'iniziativa lanciata nel 2017 dal Dipartimento della difesa americano, *Project Maven* (Manson, 2024), è l'esempio di questo cambiamento. Il sistema *Maven*, infatti, è progettato per utilizzare l'IA per processare grandi quantità di filmati da droni e satelliti, e assistere nell'identificazione dei bersagli. Un sistema simile è *Lavender*, il sistema sviluppato da Israele, che stando a quanto riportato da diverse testate è stato impiegato per generare enormi liste di obiettivi per i bombardamenti a Gaza (Abraham, 2024; Dwoskin, 2024; McKernan, Davies, 2024). Anche nella Guerra russo-ucraina vengono impiegati AI-DSS per il supporto a queste operazioni (Bondar, 2024). Tuttavia, il loro uso non è senza svantaggi. Molti ufficiali delle Forze armate statunitensi, ad esempio, hanno sollevato dubbi fondamentali sull'accuratezza e la robustezza di *Maven*. D'altro canto, l'uso di *Lavander* sembra aver favorito un approccio indiscriminato nella selezione degli obiettivi da parte di Israele, sollevando dubbi sulla sua affidabilità e sul rispetto del DIU. Da questi recenti esempi, sembra evidente il rischio di una dipendenza eccessiva dall'IA, particolarmente quando il controllo umano risulta diminuito, che può portare a complicazioni di tipo etico e legale.

2. Il rischio per il rispetto del diritto internazionale umanitario

La letteratura identifica come una delle preoccupazioni principali legate all'uso di AI-DSS quella dell'adesione alle norme di diritto umanitario internazionale (Nadibaidze *et al.*, 2024). In effetti, i sistemi di IA sono ampiamente ritenuti incapaci di interpretare correttamente i principi del DIU, in particolare quelli di proporzionalità e distinzione (Crootof, 2015; Sharkey, 2014; Szpak, 2020; Thurnher, 2018). Il principio di distinzione richiede di discriminare tra obiettivi di valore militare e non, valutando caso per caso. Il procedimento è ancora più complesso per bersagli umani (Crootof, 2015), in particolare in conflitti nei centri urbani, in cui la popolazione civile è più coinvolta e a volte partecipa direttamente alle ostilità, spesso senza un'evidente distinzione visiva tra combattenti e non (Szpak, 2020). Tuttavia, le tecnologie a disposizione delle forze armate ad oggi non permettono di compiere queste fini distinzioni.

In ogni caso, se pure la tecnologia avanzasse al punto da rendere possibile una corretta applicazione del principio di distinzione, il principio di proporzionalità sarebbe ancora un ostacolo. Il principio di proporzionalità, infatti, richiede di valutare caso per caso se il vantaggio di un'azione militare superi il danno collaterale previsto. Tuttavia, nonostante siano già in uso metodi computazionali per la valutazione della proporzionalità (ad esempio la *Collateral Damage Estimation Methodology*), questi sono pensati per fornire esclusivamente indicazioni di massima e non risultati definitivi. In più, intensificare il loro uso attraverso AI-DSS rischia di snaturare il principio di proporzionalità stesso (Gunneflo, Noll, 2023). D'altra parte, non esiste una regola fissa per confrontare il vantaggio militare di un'azione e i danni collaterali della stessa, né se ne dovrebbe derivare una da esempi passati, rendendo necessario valutare l'appropriatezza e legittimità di ciascun attacco caso per caso (Sharkey, 2014). Questo è completamente estraneo ai contemporanei sistemi di IA basati sul *machine learning* (ML), poiché essi basano la loro potenza sulla generalizzazione sulla base dei dati su cui sono addestrati.

Esistono anche problematiche di più ampio respiro. I sistemi ML sono noti per avere difficoltà nell'interpretazione dei principi giuridici in generale e nel "ragionamento giuridico" (Lai *et al.*, 2023). Inoltre, per interpretare accuratamente i principi di DIU è necessario una sorta di "giudizio umano" radicato nel buonsenso e nella buonafede (Szpak, 2020). Tuttavia, esiste un ampio consenso sulle limitate capacità di "ragionamento" dei sistemi ML (Bishop, 2021; Dentella *et al.*, 2024; Leivada *et al.*, 2024) e altrettanto sulla loro incapacità ad usare il "buonsenso" (Davis, Marcus, 2015).

3. L'automation bias e l'uso dell'IA in contesti militari

Come abbiamo visto sopra, nelle menti dei primi progettatori di AI-DSS vi era già la consapevolezza del rischio che potessero influenzare le decisioni umane ed introdurre in questo modo dei bias cognitivi. In effetti, questa si è confermata una delle problematiche principali dell'interazione uomo-macchina, compreso in campo militare (Dobbe, Wolters, 2024; Nadibaidze *et al.*, 2024). Alcuni dei primi studi in materia, effettuati nel campo dell'aviazione, hanno portato alla definizione del termine *automation bias* come gli "errori compiuti quando gli operatori umani utilizzano i suggerimenti automatizzati come un rimpiazzo euristico della raccolta ed analisi vigile delle informazioni" (Mosier *et al.*, 1998: 50-51). Tutt'oggi la maggior parte degli studi in materia si concentra in campo medico o aerospaziale (Goddard *et al.*, 2012), solo più recentemente l'impatto dell'*automation bias* è stato studiato anche nel campo della pubblica amministrazione (Ruschemeier, Hondrich, 2024; Schäferling, 2023), delle forze dell'ordine (Englezos, 2023; Selten *et al.*, 2023) e della giustizia (Gravett, 2024; Zerilli, 2022), mostrando come questo fenomeno si verifichi su ampia scala. Inoltre, l'*automation bias* può interagire con altri tipi di "scorciatoie mentali" come la *machine heuristic*, ovvero il meccanismo per cui gli umani sono portati a ritenere le macchine più "oggettive", "imparziali" e "affidabili" degli umani (Sundar, 2008; Sundar, Kim, 2019). Il che può portare a una maggiore dipendenza dall'automazione e di conseguenza aumentare l'*automation bias* (Sundar, 2020).

Nel contesto militare, l'*automation bias,* in combinazione con i limiti degli AI-DSS esaminate nei primi due paragrafi, può avere conseguenze critiche come, ad esempio, la selezione errata dei bersagli, oltre a una generale riduzione della *situational awarness* e del "coinvolgimento morale" degli operatori. Ciò nonostante, lo studio di questo fenomeno in campo militare è ancora ai primordi, anche se vi è un'ampia evidenza proveniente da altri settori. Inoltre, malgrado la attuale mancanza di dati empirici sull'impatto dell'*automation bias* in ambito militare, non è difficile immaginare come questo possa svilupparsi. Infatti, secondo la *dual process theory* di Kahneman gli umani in condizioni di stress utilizzano i processi del "sistema 1", più veloci e meno cognitivamente impegnativi, i quali tuttavia possono portare all'introduzione di bias cognitivi (Kahneman, 2011), come l'*automation bias*. Le operazioni militari sono un chiaro esempio di ambiente ad alto stress e gli ufficiali militari non sono meno suscettibili a questo genere di scorciatoie (Berejikian, Zwald, 2024; Knighton, 2004). Inoltre, sono già disponibili alcuni casi studio consolidati che dimostrano le implicazioni *dell'automation bias* in contesti operativi militari. Un caso emblematico riguarda gli incidenti

verificatisi durante l'impiego dei sistemi di difesa aerea *Patriot* nel corso dell'invasione dell'Iraq del 2003. In particolare, nel marzo di quell'anno, durante operazioni congiunte guidate dagli Stati Uniti, un sistema *Patriot* statunitense classificò erroneamente un velivolo militare britannico come un missile iracheno in arrivo. Dato il breve tempo per reagire a loro disposizione, gli operatori scelsero di aderire alle indicazioni fornite dal sistema, procedendo ad ingaggiare il velivolo. Pur avendo la possibilità di eseguire ulteriori verifiche e di disattendere l'allerta automatizzata, essi preferirono affidarsi al responso della macchina, abbattendo così erroneamente un velivolo amico. (Coco, 2024; Cummings, 2004; Stewart, Hinds, 2023). Alla luce di precedenti come questo, sembra quindi realistico il rischio che l'*automation bias* abbia un impatto significativo sulle decisioni dei comandanti militari, qualora utilizzino AI-DSS, con conseguenze potenzialmente catastrofiche sul rispetto delle norme di diritto bellico.

4. L'IA come strumento e non sostituto del comando umano

Per mitigare i rischi dell'uso di AI-DSS in ambito militare, dunque, è necessario assicurarsi un coinvolgimento umano continuativo e significativo durante tutto il ciclo di vita dei sistemi, dalla progettazione alla valutazione post-utilizzo. È necessario che gli operatori abbiano tutto il tempo e le risorse necessarie per valutare a pieno i risultati forniti dall'IA in combinazione con altre fonti. In particolare, se la velocità di elaborazione richiesta e il volume di informazioni sono eccessivi, gli operatori devono aver a disposizione chiari protocolli per compiere questa valutazione. È fondamentale riconoscere che accelerare il processo decisionale non è sempre vantaggioso e anzi può introdurre errori con conseguenze gravi, in particolare se l'ambiente operativo è complesso. Per questo bisogna stabilire limiti chiari alla velocità delle decisioni relative all'uso della forza, specialmente in contesti in cui esiste un alto rischio di danni collaterali (Nadibaidze *et al.*, 2024). Gli AI-DSS dovrebbero essere progettati per operare a una "velocità umana" anziché "velocità macchina", preservando così il ruolo del giudizio umano nelle decisioni sull'uso della forza (ICRC, 2020).

Per ridurre il rischio di *automation bias*, una strategia è l'uso di *confidence scores*, indicatori del livello di certezza di un output. Tuttavia, tali punteggi possono indurre una fiducia eccessiva negli operatori, che potrebbero non comprendere i limiti tecnici sottostanti. Per questo, le forze armate devono fornire linee guida che chiariscano che un alto livello di confidenza non equivale automaticamente alla legittimità di un bersaglio secondo il

diritto internazionale (ICRC, 2024). Gli sviluppatori devono collaborare con i decisori militari per garantire un utilizzo informato di questi sistemi (Nadibaidze *et al.*, 2024,). Strumenti come le *model cards* possono fornire panoramiche sulle capacità e limitazioni di un sistema, aiutando gli operatori a prevedere possibili errori. Tuttavia, in ambito militare, la trasparenza di tali strumenti deve essere bilanciata con la necessità di evitare che le vulnerabilità dell'IA possano essere sfruttate da avversari (ICRC, 2024).

Infine, l'addestramento del personale militare deve includere una comprensione approfondita delle limitazioni sia umane che tecnologiche, con particolare attenzione ai problemi derivanti dall'interazione tra uomini e macchine in contesti operativi diversi. I programmi di formazione dovrebbero analizzare il modo in cui gli aspetti sociali, politici, istituzionali e tecnici interagiscono nell'uso degli AI-DSS, così da aggiornare le dottrine di *targeting* per rispondere alle sfide emergenti (Klaus, 2024; Nadibaidze *et al.*, 2024).

Conclusione e prospettive future

L'adozione degli AI-DSS nelle operazioni militari pone sfide significative, in particolare per il rischio di *automation bias* e le implicazioni sul rispetto del DIU. Sebbene questo articolo abbia esaminato tali questioni alla luce della letteratura esistente, resta la necessità di studi empirici che valutino concretamente l'impatto di questi sistemi nel contesto operativo. Un limite di questo lavoro è l'assenza di dati quantitativi e qualitativi relativi all'uso effettivo degli AI-DSS in scenari militari reali. La ricerca futura potrebbe concentrarsi su studi di casi dettagliati che analizzino incidenti operativi attribuibili all'*automation bias* o alla fiducia eccessiva nelle raccomandazioni dell'IA. Un'altra area di ricerca riguarda lo sviluppo e la valutazione di strategie di mitigazione, come l'uso di interfacce progettate per ridurre il bias o programmi di addestramento specifici. È inoltre cruciale indagare l'influenza di fattori organizzativi, culturali e istituzionali nella dipendenza dagli AI-DSS, poiché le dinamiche gerarchiche e dottrinali delle forze armate possono amplificare o attenuare il fenomeno. Infine, sarebbe utile un'analisi comparativa tra diversi contesti operativi e nazionali, per comprendere come variabili tecnologiche e normative influenzino l'integrazione dell'IA nelle decisioni militari. Solo attraverso un approccio empirico sistematico sarà possibile affinare le linee guida per un utilizzo etico e responsabile degli AI-DSS, garantendo che restino strumenti di supporto e non di deresponsabilizzazione.

Riferimenti bibliografici

Abraham Y. (2024). 'Lavender': The AI machine directing Israel's bombing spree in Gaza. *+972 Magazine*. https://www.972mag.com/lavender-ai-israeli-army-gaza/

Berejikian J.D., Zwald Z. (2024). Does military experience really matter? An empirical examination of decision framing and deterrence decision-making. *Journal of Global Security Studies*, 9(3). https://doi.org/10.1093/jogss/ogae015

Bishop J.M. (2021). Artificial intelligence is stupid and causal reasoning will not fix it. *Frontiers in Psychology*, 11: 513474. https://doi.org/10.3389/fpsyg.2020.513474

Bondar K. (2024). Understanding the military AI ecosystem of Ukraine. *Center for Strategic and International Studies*. https://www.csis.org/analysis/understanding-military-ai-ecosystem-ukraine

Carpenter C. (2024). The real 'killer robots' are already here—and they're us. *World Politics Review*. https://www.worldpoliticsreview.com/killer-robots-ai-israel-gaza/

Coco A. (2024). Exploring the impact of automation bias and complacency on individual criminal responsibility for war crimes. *Journal of International Criminal Justice*, 21(5): 1077-1096. https://doi.org/10.1093/jicj/mqad034

Cook B. (2021). The future of artificial intelligence in ISR operations. *Air & Space Power Journal*, 35(Special Issue - Perspectives on JADO): 41-55.

Crootof R. (2015). The killer robots are here: Legal and policy implications. *Cardozo Law Review*, 36(5): 1837-1915.

Cummings M. (2004). Automation bias in intelligent time critical decision support systems. *AIAA 1st Intelligent Systems Technical Conference*: 1-6. https://doi.org/10.2514/6.2004-6313

Davis E., Marcus G. (2015). Commonsense reasoning and commonsense knowledge in artificial intelligence. *Communications of the ACM*, 58(9): 92-103. https://doi.org/10.1145/2701413

Dentella V., Günther F., Murphy E., Marcus G., Leivada E. (2024). Testing AI on language comprehension tasks reveals insensitivity to underlying meaning. *Scientific Reports*, 14(1): 28083. https://doi.org/10.1038/s41598-024-79531-8

Dobbe R., Wolters A. (2024). Toward sociotechnical AI: Mapping vulnerabilities for machine learning in context. *Minds and Machines*, 34(2): 12. https://doi.org/10.1007/s11023-024-09668-y

Dwoskin E. (2024). Israel built an 'AI factory' for war. It unleashed it in Gaza. *The Washington Post*. https://www.washingtonpost.com/technology/2024/12/29/ai-israel-war-gaza-idf/

Englezos E. (2023). Policing by algorithm: NSW Police's suspect target management plan. *Alternative Law Journal*, 48(1): 17-24. https://doi.org/10.1177/1037969X221147745

Goddard K., Roudsari A., Wyatt J.C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*, 19(1): 121-127. https://doi.org/10.1136/amiajnl-2011-000089

Gravett W.H. (2024). Judicial decision-making in the age of artificial intelligence. In Sousa Antunes H., Freitas P.M., Oliveira A.L., Martins Pereira C., Vaz De Sequeira E., Barreto Xavier L. (a cura di), *Multidisciplinary Perspectives on Artificial Intelligence and the Law*, vol. 58: 281-297. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-41264-6_15

Gunneflo M., Noll G. (2023). Technologies of decision support and proportionality in international humanitarian law. *Nordic Journal of International Law*, 92(1): 93-118. https://doi.org/10.1163/15718107-bja10055

Harper J. (2018). Artificial intelligence to sort through ISR data glut. *National Defense*, 102(770): 33-35.

Holsapple C.W., Whinston A.B. (1987). *Business expert systems*. Homewood (IL): Irwin.

ICRC (2020). Artificial intelligence and machine learning in armed conflict: A human-centred approach. *International Review of the Red Cross*, 102(913): 463-479. https://doi.org/10.1017/S1816383120000454

ICRC (2024). *Expert consultation report on AI and related technologies in military decision-making on the use of force in armed conflicts*. Geneva: International Committee of the Red Cross. https://www.geneva-academy.ch/joomlatools-files/docman-files/Artificial%20Intelligence%20And%20Related%20Technologies%20In%20Military%20Decision-Making.pdf

Kahneman D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Klaus M. (2024). Transcending weapon systems: The ethical challenges of AI in military decision support systems. *Humanitarian Law & Policy*. https://blogs.icrc.org/law-and-policy/2024/09/24/transcending-weapon-systems-the-ethical-challenges-of-ai-in-military-decision-support-systems/

Knighton R.J. (2004). The psychology of risk and its role in military decision-making. *Defence Studies*, 4(3): 309-334. https://doi.org/10.1080/1470243042000344786

Lai J., Gan W., Wu J., Qi Z., Yu P.S. (2023). Large language models in law: A survey. *arXiv*. https://doi.org/10.48550/arXiv.2312.03718

Leivada E., Marcus G., Günther F., Murphy E. (2024). A sentence is worth a thousand pictures: Can large language models understand hum4n l4ngu4ge and the w0rld behind w0rds? *arXiv*. https://doi.org/10.48550/arXiv.2308.00109

Manson K. (2024). AI warfare is already here. *Bloomberg*. https://www.bloomberg.com/features/2024-ai-warfare-project-maven/

McKernan B., Davies H. (2024). 'The machine did it coldly': Israel used AI to identify 37,000 Hamas targets. *The Guardian*. https://www.theguardian.com/world/2024/apr/03/israel-gaza-ai-database-hamas-airstrikes

Mosier K.L., Skitka L.J., Heers S., Burdick M. (1998). Automation bias: Decision making and performance in high-tech cockpits. *The International Journal of Aviation Psychology*, 8(1): 47-63. https://doi.org/10.1207/s15327108ijap0801_3

Nadibaidze A., Bode I., Zhang Q. (2024). AI in military decision support systems: A review of developments and debates. *Center for War Studies*.

Remus W.E., Kottemann J.E. (1986). Toward intelligent decision support systems: An artificially intelligent statistician. *MIS Quarterly*, 10(4): 403-418.

Ruschemeier H., Hondrich L.J. (2024). Automation bias in public administration – an interdisciplinary perspective from law and psychology. *Government Information Quarterly*, 41(3): 101953. https://doi.org/10.1016/j.giq.2024.101953

Schäferling S. (2023). Identifying challenges of governmental automated decision-making. In Schäferling S. (a cura di), *Governmental Automated Decision-Making and Human Rights*, vol. 62: 93-109. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-48125-3_4

Selten F., Robeer M., Grimmelikhuijsen S. (2023). 'Just like I thought': Street-level bureaucrats trust AI recommendations if they confirm their professional judgment. *Public Administration Review*, 83(2): 263-278. https://doi.org/10.1111/puar.13602

Sharkey N. (2014). Towards a principle for the human supervisory control of robot weapons. *Politica & Società*, 2: 305-324. https://doi.org/10.4476/77105

Stewart R., Hinds G. (2023). Algorithms of war: The use of artificial intelligence in decision making in armed conflict. *Humanitarian Law & Policy*. https://blogs.icrc.org/law-and-policy/2023/10/24/algorithms-of-war-use-of-artificial-intelligence-decision-making-armed-conflict/

Sundar S.S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In Metzger M.J., Flanagin A.J. (a cura di), *Digital Media, Youth, and Credibility*: 73-100. Cambridge (MA): MIT Press. https://doi.org/10.1162/dmal.9780262562324.073

Sundar S.S. (2020). Rise of machine agency: A framework for studying the psychology of human–AI interaction (HAII). *Journal of Computer-Mediated Communication*, 25(1): 74-88. https://doi.org/10.1093/jcmc/zmz026

Sundar S.S., Kim J. (2019). Machine heuristic: When we trust computers more than humans with our personal information. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*: 1-9. https://doi.org/10.1145/3290605.3300768

Szpak A. (2020). Legality of use and challenges of new technologies in warfare: The use of autonomous weapons in contemporary or future wars. *European Review*, 28(1): 118-131. https://doi.org/10.1017/S1062798719000310

Thurnher J.S. (2018). Feasible precautions in attack and autonomous weapons. In Heintschel von Heinegg W., Frau R., Singer T. (a cura di), *Dehumanization of Warfare*: 99-117. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-67266-3_6

Zerilli J. (2022). Algorithmic sentencing: Drawing lessons from human factors research. In Ryberg J., Roberts J.V. (a cura di), *Sentencing and Artificial Intelligence*: 165-183. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780197539538.003.0009

Intimità artificiale. Una prospettiva critica sull'integrazione dei Sex Robot nel lavoro sessuale

di Fabrizia Pasciuto*

L'introduzione dei Sex Robot nel lavoro sessuale solleva questioni etiche, legali e sociali. La diffusione di case di appuntamento con bambole in silicone e robot evidenzia la mancanza di regolamentazioni adeguate, esponendo gli utenti a rischi relativi alla privacy e alla sicurezza. Questo scenario invita a riflettere su come bilanciare innovazione, tutela dei diritti e implicazioni culturali di tali tecnologie.

Parole chiave: Sex Robot; lavoro sessuale; intelligenza artificiale; rischi; privacy; sicurezza.

Artificial intimacy. A critical perspective on the integration of Sex Robots into sex work

The introduction of Sex Robots into the sex work industry raises ethical, legal and social concerns. The emergence of brothels offering Sex Dolls and robots highlights the lack of adequate regulation and exposes users to potential privacy and safety risks. This scenario calls for reflection on how to balance innovation, the protection of rights and the cultural implications of such technologies.

Keywords: Sex Robot; sex work; artificial intelligence; risk; privacy; security.

Introduzione

L'intelligenza artificiale (IA) è ormai parte integrante della vita quotidiana, con effetti sempre più visibili anche in ambiti considerati intimi e personali, come le relazioni affettive e la sessualità. In questo contesto, si collocano le tecnologie del cosiddetto *sextech*, e in particolare i Sex Robot, dispositivi progettati per simulare esperienze intime personalizzate, che aprono interrogativi non solo tecnici, ma profondamente culturali e politici (Jin, Pena, 2010; McDaniel, Coyne 2016; Parsakia, Rostami 2023).

Oltre agli utilizzi ludici, queste tecnologie sono state proposte come strumenti terapeutici per persone disabili (Pasciuto, Cava, Falzone 2023), per il benessere degli anziani (Jecker, 2021), per la riabilitazione dei sex offenders

DOI: 10.5281/zenodo.17524696

* Università della Tuscia. fpasciuto@unime.it.

(Zara, Veggi, Farrington 2022) o per impieghi in contesti estremi (Balistreri, 2023). Tuttavia, nessuna di queste applicazioni ha trovato finora realizzazione concreta se si esclude il settore del sex work, dove i Sex Robot sono già offerti come servizio in vere e proprie case di appuntamento dedicate ad essi.

Questa sperimentazione non solleva solo questioni morali o normative, ma mette anche in luce le implicazioni legate alla raccolta, gestione e monetizzazione dei dati sensibili. Infatti, come è stato osservato (Zuboff, 2023), molte tecnologie digitali, pur presentandosi come strumenti di libertà e personalizzazione, funzionano in realtà secondo logiche economiche basate sulla sorveglianza e sullo sfruttamento dell'informazione privata. Quando questo modello viene applicato alla sessualità, il rischio non è solo la perdita di privacy, ma la trasformazione stessa dell'intimità in un asset commerciale.

La crescente diffusione di dispositivi e applicazioni che raccolgono dati intimi, dalle app per il monitoraggio della salute sessuale (Gross *et al.*, 2021) a quelle di dating (Phan *et al*., 2021), fino ai sex toys e ai Sex Robot, ha già sollevato casi concreti di violazione della privacy (Kindt, 2013). Aziende produttrici di articoli per il benessere sessuale come *Lovense* o *We-Vibe* sono state accusate di immagazzinare dati sensibili degli utenti senza consenso esplicito (Sundén, 2023; Stardust, 2024), mostrando come la promessa di esperienze personalizzate si accompagni spesso a dinamiche opache e invasive. In questo scenario, il corpo sessuato diventa, come osserva Lupton (2016), una fonte costante di dati da elaborare, archiviare e monetizzare.

I modelli più avanzati di Sex Robot non si limitano a simulare movimenti o conversazioni in maniera pre-programmata: grazie all'IA integrata, a sensori disseminati sotto la pelle sintetica e app dedicate, possono registrare stimoli tattili, preferenze individuali, reazioni fisiologiche e persino stati emozionali. La casa produttrice più celebre per la creazione di Sex Robot, la statunitense *Abyss Creations*, ha attualmente in fase di sviluppo anche una tecnologia che consentirebbe l'integrazione di telecamere nei loro occhi, facendo nascere ulteriori dubbi riguardo alla possibilità che diverranno in grado di registrare e archiviare filmati intimi degli utenti. Questo rende l'esperienza altamente personalizzata, ma introduce interrogativi sulla raccolta automatica di dati estremamente intimi. Il confine tra corpo e mercato, tra esperienza privata e profilazione algoritmica, si fa sempre più labile. Per questo motivo, i Sex Robot non possono essere considerati soltanto innovazioni nel campo del sextech: sono artefatti che condensano visioni del corpo, del genere, del desiderio e del potere. Questo lavoro si inserisce negli studi critici sulla tecnologia, adottando una prospettiva sociologica che interpreta tali dispositivi come prodotti culturali e sociali, che riflettono, e

potenzialmente rinforzano, disuguaglianze esistenti, dinamiche di controllo e logiche di consumo (Fuchs, 2017; Haraway, 1991; Jasanoff, 2004).

Il focus sul sex work nasce dalla concreta sperimentazione dei Sex Robot in questo ambito: nonostante i costi elevati (oltre 20.000 dollari per modello), sono già utilizzati in contesti reali, dove sollevano interrogativi normativi, etici e politici. L'intreccio tra Sex Robot e sex work può essere visto in una prospettiva multiforme. Da un lato, potrebbe essere considerato una soluzione al contrasto del traffico di esseri umani e alle malattie sessualmente trasmissibili. Dall'altro, dobbiamo considerare i diritti di chi ha scelto liberamente di lavorare nel sex work, ora potenzialmente minacciato da questi artefatti. Pertanto, la questione non riguarda solo l'utente o il dispositivo, ma chi, nel sex work tradizionale, rischia di essere reso invisibile o sostituibile. L'automazione dell'intimità, quando coinvolge questo settore, richiede una riflessione che tenga insieme corpo, consenso, dati e disuguaglianza.

1. Sex Robot/Dolls Brothels in Europa e nel mondo

Nonostante nessun Paese disponga ancora di un quadro normativo chiaro sull'impiego dei Sex Robot, il loro utilizzo solleva interrogativi etici e giuridici, soprattutto quando assume forme che imitano comportamenti socialmente sensibili: basti pensare ai casi, già oggetto di divieti morali, in cui i robot assumono sembianze infantili (Maras, Shapiro, 2017) o vengono impiegati nel settore della prostituzione (Marchant, Climbingbear, 2022).

Alcuni dei primi esempi di impiego commerciale delle Sex Dolls – bambole in silicone dalle sembianze iperrealistiche – risalgono al Giappone e alla Corea del Sud, dove già a partire dai primi anni Duemila si sono diffusi servizi di escort basati sull'affitto di questi prodotti. In Europa, il fenomeno ha iniziato a prendere piede con l'apertura, nel 2017, del bordello *Lumidolls* a Barcellona, a cui sono seguite sedi anche in Italia e Germania e in altri Paesi europei e del Nord America (Xuan, 2019). Nonostante il successo iniziale, molti di questi esercizi sono stati successivamente chiusi a causa dell'assenza di normative adeguate che ne regolamentassero le attività.

Al di là della dimensione imprenditoriale, questi bordelli rappresentano un campo di osservazione utile per comprendere i mutamenti nella concezione dell'intimità, della sessualità e del lavoro. Alcune testimonianze raccolte da media internazionali hanno evidenziato come parte dei clienti prediliga l'assenza di interazione verbale e di reciprocità, percependo la relazione con la bambola come uno spazio privo di giudizio e conflitto (BBC 2019). Questa dinamica sembra rispondere a un modello di intimità

semplificata, dove il desiderio viene soddisfatto in assenza di alterità e negoziazione.

In una prospettiva sociologica, l'esperienza di questo particolare tipo di case di appuntamento può essere letta come una forma estrema di oggettificazione del corpo (Nussbaum, 1995), in cui il partner sessuale è interamente adattabile e gestibile, ridotto a superficie programmabile. Ma non si tratta soltanto di una riproduzione delle dinamiche di dominio sul corpo femminile: queste tecnologie incarnano una trasformazione più profonda, in cui l'intimità stessa viene riformulata secondo logiche prestazionali. Seguendo le riflessioni di Bauman (2013) e Illouz (2007), si potrebbe parlare di una "intimità senza attrito", dove la relazione è svuotata di incertezza, conflitto, reciprocità, e resa pienamente funzionale al consumo individuale. I Sex Robot si presentano così come strumenti di semplificazione del legame, rispondenti a una logica neoliberale che riduce la sessualità a servizio personalizzato e replicabile. In questa visione, il Sex Robot non è solo una tecnologia nell'ambito del mercato del benessere sessuale, ma un artefatto sociale che veicola una precisa idea di relazione: immediata, unidirezionale, priva di negoziazione e orientata all'efficienza. Questa lettura permette di cogliere come la crescente automazione dell'intimità non elimini i rapporti di potere, ma li riproduca sotto nuove forme, spesso più difficili da riconoscere perché mediate da dispositivi all'apparenza neutrali. In realtà, il Sex Robot è portatore di valori culturali ben definiti, e la sua progettazione riflette, nella maggior parte dei casi, un'idea stereotipata di desiderio maschile eterosessuale.

Nel dibattito internazionale, studiosi e studiose hanno assunto posizioni divergenti in merito alla diffusione di questi artefatti. Levy (2007; 2009) ritiene che i Sex Robot possano ridurre i rischi legati alla prostituzione tradizionale, offrendo esperienze intime più sicure. All'opposto si colloca Richardson (2016), la quale sostiene che tali dispositivi rischiano di disumanizzare la sessualità e normalizzare dinamiche relazionali distorte. In una prospettiva radicale, Bryson (2010) ha persino ipotizzato che, non essendo soggetti umani, i robot potrebbero assumere il ruolo di "schiavi eticamente accettabili", soggetti a coercizioni senza vittime reali.

Anche nel contesto italiano il tema dei Sex Robot ha iniziato a suscitare attenzione, dando luogo a un confronto tra approcci giuridici, etico-filosofici e riflessioni femministe. Tra le prime voci a dare rilevanza a questo dibattito si colloca quella di Balistreri (2018), che affronta l'argomento rifiutando una lettura moralistica dei gusti sessuali degli individui. Secondo l'autore, l'uso dei Sex Robot non può essere condannato in sé, né associato automaticamente a un aumento delle violenze o dei comportamenti sessuali devianti. Per sostenere questa posizione, l'autore richiama diversi studi sul rapporto

tra contenuti mediali violenti – come i videogiochi – e comportamenti degli utenti che vi entrano in contatto, sottolineando come il loro consumo non implichi necessariamente una ricaduta nella realtà. Su una linea affine si muove Marrone (2018), che adotta una prospettiva centrata sull'ibridazione tra umano e macchina. Attraverso la valorizzazione della natura tecnica dell'essere umano, lo studioso sostiene che tali tecnologie non rappresentino una rottura, ma la continuazione di un processo storico di integrazione uomo-tecnica. L'approccio di Balistreri viene qui condiviso e definito "pragmatico" e "antiantropocentrico", in quanto capace di sfuggire tanto ai moralismi quanto a visioni apocalittiche dell'innovazione.

Di segno diverso è invece la posizione di Rigotti (2020), che ne propone una lettura femminista e intersezionale. L'autrice mette in discussione l'idea che i Sex Robot possano essere strumenti di emancipazione, evidenziando come la loro progettazione e commercializzazione siano ancora fortemente segnate da logiche androcentriche e da una sessualità normata secondo standard maschili ed eterosessuali. Perché queste tecnologie possano avere un potenziale trasformativo, Rigotti sostiene la necessità di aprire il design e la produzione a soggettività femminili e queer, promuovendo modelli alternativi di desiderio, piacere e rappresentazione dei corpi.

Il confronto tra queste prospettive solleva una questione centrale che riguarda non solo la sessualità, ma anche il lavoro: cosa accade quando il sex work viene disumanizzato attraverso l'introduzione di macchine? I Sex Robot, in questo senso, non si limitano a offrire un'alternativa tecnologica alla prestazione sessuale, ma rischiano di ridefinire i termini stessi della relazione tra cliente e lavoratore. Laddove il sex work, oggi, implica, in misura variabile, elementi di agency, negoziazione, vulnerabilità e soggettività, l'interazione con un robot cancella ogni margine di reciprocità, rendendo l'esperienza completamente centrata sul consumo individuale.

Questo slittamento ha conseguenze dirette sul riconoscimento culturale e politico del sex work come lavoro. Se il Sex Robot viene percepito come soluzione "etica" e "neutra" al lavoro sessuale, allora il sex work tradizionale rischia di essere ulteriormente delegittimato, letto come qualcosa di superabile e non come una forma di attività meritevole di tutela. In quest'ottica, i Sex Robot non sono affatto neutrali, ma partecipano attivamente alla costruzione simbolica di cosa è o non è accettabile in termini di sessualità e lavoro.

2. Scenari possibili nella regolamentazione dei Sex Robot

Di fronte a questa complessità si delineano alcuni possibili scenari, ciascuno dei quali riflette priorità e visioni differenti. Una regolamentazione non è, quindi, una questione meramente tecnica ma coinvolge i tentativi di stabilire se e come bilanciare l'innovazione con la protezione dei diritti individuali e collettivi (Lev 2022). Tra le opzioni attualmente ipotizzabili, si possono delineare quattro approcci principali:

- *Non regolamentare i Sex Robot*: questo approccio laissez-faire, legato al postulato della libertà individuale, presupporrebbe che i Sex Robot siano trattati come un prodotto di consumo qualsiasi, lasciando al mercato la responsabilità di definirne gli standard e le pratiche. Tuttavia, in assenza di regolamentazione gli utenti potrebbero essere esposti a prodotti potenzialmente non sicuri e il trattamento dei dati personali rimarrebbe largamente non controllato.
- *Integrare i Sex Robot nelle leggi vigenti*: i Sex Robot dovrebbero seguire sia le indicazioni relative ai prodotti per il benessere sessuale, sia quelle che regolano il lavoro sessuale nelle legislazioni dei diversi Paesi. Sarebbero quindi soggetti agli stessi standard di sicurezza e qualità richiesti per i sex toys, con l'aggiunta di regolamentazioni specifiche sul lavoro sessuale laddove pertinente. Tuttavia, Sex Dolls e Sex Robot sono stati spesso utilizzati per aggirare le leggi esistenti sfruttando vuoti normativi. Questo approccio potrebbe offrire una cornice legale già consolidata, ma potrebbe risultare insufficiente a catturarne le peculiarità tecnologiche e le implicazioni sociali.
- *Creare leggi apposite per la regolamentazione dei Sex Robot*: una soluzione più strutturata consisterebbe nello sviluppo di un quadro normativo specifico con disposizioni mirate alla sicurezza dei materiali, alla tutela dei dati personali e alla regolamentazione in contesti commerciali. Tale approccio riconoscerebbe la natura unica di questa tecnologia, ma comporterebbe uno sforzo legislativo significativo e il rischio di una lenta implementazione. Inoltre, in uno scenario internazionale frammentato alcuni stati potrebbero optare per un uso più liberale, mentre altri potrebbero imporre restrizioni o divieti basati su influenze culturali, morali o religiose (Szczuka, Krämer 2017; Oleksy, Wnuk 2021; Karaian, 2022; Brandon, Shlykova, Morgentaler, 2022).
- *Vietare del tutto i Sex Robot*: una posizione radicale potrebbe consistere in un divieto totale sulla produzione, la distribuzione e

l'utilizzo dei Sex Robot. Questo approccio è sostenuto da Kathleen Richardson, la quale già nel 2015 ha fondato la "*Caimpaign Against Sex Robots*" (oggi "*Campaign Against Porn Robots*") motivata da preoccupazioni come il rischio di disumanizzazione del prossimo e l'aumento delle violenze sulle donne trattate come meri oggetti sessuali (Richardson, 2016; 2018; 2023). Questa soluzione potrebbe essere difficile da mettere in atto sia a causa della natura globale del mercato tecnologico e della crescente domanda di dispositivi basati sull'IA, ma anche perché l'applicazione di un divieto totale richiederebbe un ampio consenso internazionale, difficilmente raggiungibile.

Come già evidenziato, il mercato dei Sex Robot riflette e amplifica gli squilibri di genere già presenti nella società. Se osserviamo la predominanza di donne cisgender e transgender nella prostituzione e la prevalenza di uomini tra i clienti (Smith, Mac 2018), possiamo notare come il mercato dei Sex Robot sembra riproporre queste stesse dinamiche, con una forte presenza di modelli femminili che incarnano un'idea specifica di desiderabilità e sessualizzazione per un pubblico prevalentemente maschile (Pasciuto, 2024). Infatti, sebbene alcune case produttrici abbiano accessori opzionali come il *transgender converter*, che permette di modificare l'anatomia del robot per includere un organo genitale maschile, questo elemento sembra più un adattamento accessorio che una reale apertura alla diversità di genere e orientamenti sessuali.

Pertanto, è possibile affermare che la regolamentazione dei Sex Robot non può essere intesa come un semplice problema tecnico, ma come un campo di contesa tra visioni diverse della sessualità, del corpo e della cittadinanza (Supiot, 2005). Anche l'assenza di norme, il cosiddetto approccio laissez-faire, è in sé una forma di governance che lascia spazio ad una regolazione opaca (Yeung, 2018), spesso guidata da interessi di mercato.

Riflessioni conclusive

L'introduzione dei Sex Robot nel lavoro sessuale non rappresenta soltanto un'innovazione tecnologica, ma una trasformazione simbolica che investe le modalità con cui la società costruisce e regola l'intimità, il desiderio e la corporeità. Lungi dall'essere strumenti neutri, questi dispositivi incorporano modelli culturali di genere, sessualità e potere, riproducendo – spesso in modo acritico – un immaginario dominato dalla sessualizzazione del corpo femminile e dalla sua piena disponibilità al consumo maschile. Il loro

impiego nel sex work, sebbene ancora limitato, segnala una tendenza più ampia alla mercificazione dell'intimità e alla sostituzione dell'interazione relazionale con forme prestazionali automatizzate.

Come si è visto, il fenomeno solleva questioni che attraversano diversi ambiti disciplinari: dal diritto, chiamato a confrontarsi con l'assenza di un quadro normativo specifico, all'etica, passando per le scienze sociali, che devono interrogarsi sulle conseguenze culturali e materiali di queste tecnologie. In particolare, l'introduzione dei Sex Robot rischia di produrre effetti ambivalenti: se da un lato promette esperienze più sicure, dall'altro può contribuire alla delegittimazione del sex work umano e rafforzare narrazioni stigmatizzanti nei confronti dei lavoratori e delle lavoratrici del sesso.

Il caso dei bordelli con bambole e robot sessuali mostra come queste tecnologie, più che ridurre la complessità delle relazioni, tendano a semplificarla fino ad annullarne le dimensioni di reciprocità, negoziazione e imprevedibilità. In questo senso, i Sex Robot non modificano soltanto l'offerta di sex work, ma interrogano i fondamenti stessi delle relazioni interpersonali nell'epoca della pervasività tecnologica.

Più che risolvere problemi sociali preesistenti, queste tecnologie rischiano di consolidare disuguaglianze, marginalizzazioni e visioni stereotipate del desiderio maschile e del corpo femminile. Interrogarle criticamente non significa temere l'innovazione, ma riconoscere che ogni scelta tecnica è anche una scelta culturale, politica e simbolica.

In definitiva, i Sex Robot rappresentano una delle manifestazioni più estreme della logica neoliberale applicata alla sfera dell'intimità: una logica che frammenta le relazioni, riduce la sessualità a prestazione misurabile, elimina l'alterità a favore della prevedibilità. L'intimità diventa servizio, il corpo diventa piattaforma, il desiderio diventa algoritmo. In questa prospettiva, il Sex Robot non è solo un dispositivo tecnologico, ma un prodotto perfettamente allineato alla razionalità capitalista contemporanea, che trasforma ogni esperienza affettiva in occasione di consumo e ogni soggettività in oggetto di profilazione.

Il punto, probabilmente, non è essere *a favore* o *contro* i Sex Robot, ma interrogarsi su quali visioni della sessualità, del corpo e del lavoro essi incorporano, e su quali soggettività rischiamo di lasciare fuori dal discorso. Una riflessione critica non può limitarsi a registrare l'innovazione, ma deve chiedersi chi ne trae giovamento o profitto, chi li regola e chi viene escluso.

Riferimenti bibliografici

Balistreri M. (2018). *Sex robot: l'amore al tempo delle macchine*. Roma: Fandango.

Balistreri M. (2023). Le questioni morali e le implicazioni psicologiche della riproduzione, del sesso e delle relazioni affettive nelle missioni spaziali. *Rivista internazionale di Filosofia e Psicologia*, 14(3): 148-167.

Bauman Z. (2013). *Liquid love: On the frailty of human bonds*. London: John Wiley & Sons.

BBC News (2019). *Sex Doll Brothels: A Growing Trend?* BBC News. https://www.youtube.com/watch?v=pTSrLHxSoAQ (consultato il 20 marzo 2024).

Brandon M., Shlykova N., Morgentaler A. (2022). Curiosity and other attitudes towards sex robots: Results of an online survey. *Journal of Future Robot Life*, 3(1): 3-16.

Bryson J.J. (2010). Robots should be slaves. In *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*, 8: 63-74.

Fuchs C. (2017). *Social media: A critical introduction* (2ª ed.). London: Sage.

Gross M.S., Hood A., Corbin B. (2021). Pay no attention to that man behind the curtain: An ethical analysis of the monetization of menstruation app data. *International Journal of Feminist Approaches to Bioethics*, 14(2): 144-156.

Haraway D. (1991). A cyborg manifesto: Science, technology, and socialist-feminism in the late twentieth century. In *Simians, Cyborgs and Women: The Reinvention of Nature*: 149-181. New York: Routledge.

Illouz E. (2007). *Cold intimacies: The making of emotional capitalism*. Cambridge: Polity.

Jasanoff S. (2004). *States of knowledge*. Abingdon (UK): Taylor & Francis.

Jecker N.S. (2021). Nothing to be ashamed of: Sex robots for older adults with disabilities. *Journal of Medical Ethics*, 47(1): 26-32.

Jin B., Pena J.F. (2010). Mobile communication in romantic relationships: Mobile phone use, relational uncertainty, love, commitment, and attachment styles. *Communication Reports*, 23(1): 39-51.

Karaian L. (2022). Plastic fantastic: Sex robots and/as sexual fantasy. *Sexualities*, 0(0): 1-20.

Kindt E.J. (2013). *Privacy and data protection issues of biometric applications: A comparative legal analysis*, vol. 12. Cham: Springer.

Lev D. (2022). These are not the droids you are looking for: The urgent need for state regulation of artificially intelligent sex robots. *Illinois Journal of Law, Technology & Policy*: 483.

Levy D. (2007). Robot prostitutes as alternatives to human sex workers. In *IEEE International Conference on Robotics and Automation*, vol. 14. Rome.

Levy D. (2009). *Love and sex with robots: The evolution of human-robot relationships*. New York.

Lupton D. (2016). *The quantified self*. London: John Wiley & Sons.

Maras M.H., Shapiro L.R. (2017). Child sex dolls and robots: More than just an uncanny valley. *Journal of Internet Law*, 21(5): 3-21.

Marchant G.E., Climbingbear K. (2022). Legal resistance to sex robots. *Journal of Future Robot Life*, 3(1): 91-107.

Marrone P. (2018). Lovotics: técnica, natura, sex robot. *Diritto & Questioni Pubbliche*, 18(2): 239-250.

McDaniel B.T., Coyne S.M. (2016). "Technoference": The interference of technology in couple relationships and implications for women's personal and relational well-being. *Psychology of Popular Media Culture*, 5(1): 85.

Nussbaum M.C. (1995). Objectification. *Philosophy & Public Affairs*, 24(4): 249-291.

Oleksy T., Wnuk A. (2021). Do women perceive sex robots as threatening? The role of political views and presenting the robot as a female-vs male-friendly product. *Computers in Human Behavior*, 117: 106664.

Parsakia K., Rostami M. (2023). Digital intimacy: How technology shapes friendships and romantic relationships. *AI and Tech in Behavioral and Social Sciences*, 1(1): 27-34.

Pasciuto F. (2024). Sessualità e tecnologia: La rappresentazione del corpo femminile nella costruzione dei sex robot. In *Gender R-Evolutions: immaginare l'inevitabile, sovvertire l'impossibile*, vol. 8: 229-240. Trento: Università degli Studi di Trento.

Pasciuto F., Cava A., Falzone A. (2023). The potential use of sex robots in adults with autistic spectrum disorders: A theoretical framework. *Brain Sciences*, 13(6): 954.

Phan A., Seigfried-Spellar K., Choo K.K.R. (2021). Threaten me softly: A review of potential dating app risks. *Computers in Human Behavior Reports*, 3: 100055.

Richardson K. (2016). The asymmetrical 'relationship' parallels between prostitution and the development of sex robots. *ACM SIGCAS Computers and Society*, 45(3): 290-293.

Richardson K. (2018). *Sex robots: The end of love*. Cambridge: Polity Press.

Richardson K. (2023). The end of sex robots: Porn robots and representational technologies of women and girls. In *Man-Made Women: The Sexual Politics of Sex Dolls and Sex Robots*: 171-192. Cham: Springer International Publishing.

Rigotti C. (2020). Guardare i sex robots attraverso le lenti femministe. *Filosofia*, (65): 21-38.

Stardust Z. (2024). Sex tech in an age of surveillance capitalism: Design, data and governance. In *Routledge Handbook of Sexuality, Gender, Health and Rights*: 448-458. London: Routledge.

Sundén J. (2023). Play, secrecy and consent: Theorizing privacy breaches and sensitive data in the world of networked sex toys. *Sexualities*, 26(8): 926-940.

Supiot A. (2005). *Homo juridicus: Essai sur la fonction anthropologique du droit*. Paris: Éditions du Seuil.

Szczuka J.M., Krämer N.C. (2017). Not only the lonely—How men explicitly and implicitly evaluate the attractiveness of sex robots in comparison to the attractiveness of women, and personal characteristics influencing this evaluation. *Multimodal Technologies and Interaction*, 1(1): 3.

Xuan P.T.H. (2019). From the sex doll in the doll hotel in the 2018 World Cup season: The globalization context. In *16th International Symposium on Management (INSYMA 2019)*: 1-4.

Yeung K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance*, 12(4): 505-523.

Zara G., Veggi S., Farrington D.P. (2022). Sexbots as synthetic companions: Comparing attitudes of official sex offenders and non-offenders. *International Journal of Social Robotics*, 14(2): 479-498.

Zuboff S. (2023). The age of surveillance capitalism. In *Social Theory Re-wired*: 203-213. London: Routledge.

Chatbot, antropomorfizzazione e intelligenza artificiale: una sfida formativa

di Davide Pedone*

L'intelligenza artificiale rivoluziona ogni ambito lavorativo, dalle attività manuali a quelle intellettuali. Questo contributo esamina il fenomeno, evidenziando il ruolo dell'alfabetizzazione nell'affrontare sfide come l'antropomorfizzazione. Un approccio interdisciplinare, che unisca diverse aree del sapere, è fondamentale per comprendere e governare l'impatto dell'IA sulla società.

Parole chiave: uomo; rivoluzione; IA; educazione; etica; antropomorfizzazione.

Chatbot, anthropomorfization and artificial intelligence: an educational challenge

AI is reshaping every work sector, from manual to intellectual tasks. This study explores the phenomenon, emphasizing the importance of literacy in tackling challenges like anthropomorphization. A multidisciplinary approach, merging different fields of knowledge, is crucial to grasping and managing AI's societal impact.

Keywords: human; revolution; IA; education; ethic; anthropomorphization.

1. IA e sociologia del rischio: uno sguardo introduttivo

Le indagini sociologiche intorno al mondo tecnologico trovano un loro sviluppo a partire dagli anni Sessanta e Settanta del Novecento, quando si sono iniziati a vedere i primi effetti che venivano a realizzarsi. Questo interesse si muove per rilevare non solo l'influenza dei vari artefatti tecnologici ma capire come riuscissero a modellare profondamente la società, indagandone usi ed abusi (Gobo, Marcheselli, 2021). Tali indagini non erano presenti nel passato in quanto si riteneva che gli strumenti tecnologici fossero legati al processo innovativo, cioè il motore della società (Gobo, Marcheselli, 2021).

In realtà, la tecnologia come ricorda Melvin Kranzberg: «non è né buona, ne cattiva né neutrale» (Kranzeberg, 1986: 547). Oggi, è chiaro che la tecnologia ha un potere trasformativo sulla società, influenzando usi, costumi e

DOI: 10.5281/zenodo.17524714

* Università degli Studi G. D'Annunzio di Chieti-Pescara. davide_ped99@outlook.it.

dinamiche sociali in modo visibile e profondo (Rocca, 2024). Questo risulta ancora più evidente nella società odierna con lo sviluppo dei sistemi intelligenti.

L'emergere dell'intelligenza artificiale, divenuta oggi tecnologia dominante e di spiccato interesse, può trovare una cornice interpretativa all'interno delle teorie sociologiche del rischio. L'IA incarna tutti gli elementi tipici della modernità riflessiva: globalità, invisibilità, anticipazione e produzione endogena (Beck, 1986). La modernità, come descritta da Beck, si caratterizza per «ascrittività» (Beck, 2000: 54), elemento attraverso cui l'autore vuole evidenziare come l'uomo, grazie al progresso tecnologico e scientifico, produce dei manufatti che, sfuggendo al controllo, hanno la potenzialità di generare incertezze sistemiche e globali. Questi elementi emergono in un'intervista dove il sociologo sottolinea che la rapidità dello sviluppo tecnologico ha prodotto un incremento di rischi e insicurezza, generando delle conseguenze imprevedibili (Yates, 2016).

Manuel Castells, nel suo libro *La nascita delle società in rete* (1996), anticipa come le nuove tecnologie possano ristrutturare le relazioni e le strutture sociali, un fenomeno amplificato dall'avvento dell'IA. Le nuove tecnologie impongono una necessaria riconfigurazione degli assetti strutturali e delle dinamiche relazionali nella sfera sociale (Castells, 2014). Tale constatazione consente di porre in evidenza la natura bidirezionale dell'influenza tra dimensione sociale e apparato tecnologico, caratterizzata da un articolato meccanismo di reciprocità interattiva (Castells, 2014).

L'IA, dunque, si configura non solo come uno strumento di progresso ma anche elemento generatore di quella che Beck definirebbe nuova fase della società del rischio, dove le minacce derivano proprio da quegli artefatti tecnologici che l'umanità ha realizzato per la risoluzione dei problemi. Tali riflessioni tra società del rischio e intelligenza artificiale permettono di mettere in evidenza i benefici ma anche di mettere in risalto le rinnovate vulnerabilità. In relazione all'ultimo elemento, l'IA può mettere in evidenza e rimarcare disuguaglianze che possono essere generate dai bias algoritmici (Floridi, 2023; Aragona, 2020); può acuire il fenomeno della disinformazione, in quanto le tecnologie di IA possono contribuire alla contaminazione dello spazio informativo attraverso i *deep fake* (Filimowicz, 2022); può far emergere il fenomeno dell'antropomorfizzazione, cioè l'erronea attribuzione di peculiarità e caratteristiche umane ai sistemi di IA, sollevando interrogativi etici e sociali (Abdala Moreira, 2023). Queste sono alcune delle diverse tipologie di rischio che i nuovi sistemi tecnologici pongono all'interno della società.

In sintesi, l'intelligenza artificiale non solo rinnova ma amplifica i rischi e le vulnerabilità già esistenti, rendendo necessarie riflessioni sociologiche sull'impatto sociale che questi sistemi hanno e potranno avere nel prossimo futuro (Bennato, 2024; Crawford, 2021; Bostrom, 2023).

2. L'alfabetizzazione digitale come chiave di risoluzione

Nell'era della democratizzazione degli strumenti digitali, obiettivo dichiarato da Microsoft (Microsoft News Center, 2016), è d'obbligo analizzare la portata della rivoluzione tecnologica per cercare di sviluppare un percorso educativo che abbia come obiettivo quello di orientare gli utenti verso un utilizzo etico e consapevole delle varie tecnologie, soprattutto quelle di IA.
L'elemento chiave è l'alfabetizzazione digitale, la quale viene definita come la capacità di usare, gestire, valutare e comprendere la tecnologia (ITEA, 2007), tale processo richiede l'insegnamento e la comprensione delle nozioni di base dell'intelligenza artificiale, come funziona, cosa sia qual è l'impatto etico e sociale di tali sistemi (Ranieri, 2024). Questo è fondamentale per poter strutturare un'alfabetizzazione critica (Scarano, Ferrantino, 2024). Per giungere ad un grado corretto di conoscenza digitale, come osserva Williams, occorre strutturarla in tre dimensioni: abilità tecniche, ovvero la capacità di usare gli artefatti tecnologici; conoscenza e comprensione, che consiste nel sapere come e perché funziona; consapevolezza sociale (Williams, 2009). In merito alla strutturazione un framework di alfabetizzazione tecnologica, degna di menzione è la proposta di Stople e Hallstörm, che hanno strutturato il framework in tre step: conoscenza scientifica della tecnologia, la conoscenza delle definizioni e il funzionamento e linguaggio computazionale dei sistemi di large language model (LLM); sviluppo di abilità tecniche, padronanza dei dati, della programmazione; comprensione socio-etica, all'impatto dell'IA sulla società in termini di privacy, bias algoritmici e ripercussioni etiche (Stople, Hallstörm, 2024). L'alfabetizzazione all'intelligenza artificiale assume un ruolo centrale nel panorama formativo contemporaneo, in grado di fornire agli utilizzatori strumenti necessari per comprendere le tecnologie impiegate, il loro funzionamento algoritmico e sviluppare modalità ottimali di interazione con i sistemi. Un approccio sistematico ed interdisciplinare non solo permette di prevenire utilizzi inappropriati e disfunzionali, ma trasforma questi strumenti in risorse strategiche e utili in tutti i campi (Ciasullo, 2024). L'utilità di tali sistemi è stata evidente in molti ambiti come l'ambito chimico-medico, il campo agricolo e il campo educativo (Cfr. Limna *et al.*, 2023, Ait Baha *et al*., 2024).

Nel settore chimico-medico si può annoverare la scoperta della molecola halicin, un super antibiotico che ha la capacità di superare l'antibioticoresistenza di molti batteri, individuata grazie al sistema di intelligenza artificiale adoperato dal MIT (Kissinger *et al.*, 2024). Un ulteriore progresso nell'ambito medico è quello relativo alla scoperta e alla visualizzazione dell'avvolgimento delle proteine (Triveri, De Vivo, 2024). Questa scoperta è stata rivoluzionaria nel campo medico poiché la conoscenza del loro ripiegamento ha permesso l'accelerazione di diagnosi e la ricerca di una cura per molte malattie, riuscendo a adoperare dei trattamenti specifici per casi complessi. Entrambe le scoperte hanno portato ad un progresso epocale nel campo medico e nella ricerca e, senza gli algoritmi di intelligenza artificiale, sarebbero costate molto in termini di tempo e di risorse per la comunità scientifica.

Esempi relativi all'implementazione efficiente dei sistemi di intelligenza artificiale sono riscontrabili anche in molti settori lavorativi come quello agricolo. Negli ultimi anni, tale settore, ha subito dei danni irreparabili a causa di numerosi fattori, tra cui il cambiamento climatico. Un caso connesso è il fenomeno delle malattie che hanno colpito intere filiere agricole, fenomeno impattante e molto pervasivo. Le soluzioni che sono state adottate sono principalmente associabili a sistemi "intelligenti" in grado di analizzare, attraverso dei sensori specifici, le soluzioni adeguate per far fronte a problematiche come lo scarso stato di irrigazione della terra, cercando di automatizzarlo, o il verificare la presenza di agenti che minano il buono stato del raccolto in modo da poter somministrare, in dosi adeguate e non nocive, veleni che tutelino il raccolto, attraverso macchinari sofisticati in grado di calcolarne accuratamente il dosaggio e somministrarlo ad una determinata tipologia di prodotto agricolo (Kamilaris, Prenafeta-Boldú, 2018).

Pertanto, l'introduzione di questi sistemi integrati ha permesso di ottimizzare la filiera agricola, prevenendo i rischi ed aumentando la produttività.

Come anticipato, nonostante le numerose potenzialità che tali sistemi riservano, è necessario e doveroso concentrarsi sui rischi associati. Come evidenziato nello studio empirico di Hyeon Jo, alcuni elementi che possono inibire l'uso dei sistemi di LLM possono essere rintracciati nella privacy, minando la fiducia che l'utilizzatore potrebbe avere nei confronti di tali sistemi, nella tecnofobia, rintracciata principalmente negli utenti meno esperti o, nel caso specifico degli studenti, nel senso di colpa poiché viene percepita come una scorciatoia, una forma di imbroglio (Jo, 2024). Ulteriori rischi vengono evidenziati nello studio di Tarchi e colleghi (2024) i quali sottolineano, oltre alle problematiche esaminate da Hyeon Jo, problemi legati al copyright o il problema delle risposte fuorvianti che possono essere fornite dai ChatBot utilizzati (Tarchi *et al.*, 2024).

Per poter arginare, prevedere ed affrontare i numerosi rischi associati ai sistemi di intelligenza artificiale, risulta necessario comprenderne il funzionamento. L'alfabetizzazione risulta essere il fondamento imprescindibile per poter affrontare tali implicazioni tecnologiche, sociali ed etiche dell'IA. Dunque, è necessario insegnare, da un punto di vista pratico, l'utilizzo di tali sistemi negli ambienti scolastici (Jo, 2024) affinché, in una società sempre più tecnologizzata, possano essere strumenti benefici concreti in tutti gli ambiti: educativo, lavorativo e quotidiano.

In ambito educativo, un beneficio è stato osservato nello studio di Zhang e Huang (2024) che, attraverso una ricerca empirica robusta che ha combinato analisi qualitative e quantitative, ha analizzato i benefici dei ChatBot. I risultati della ricerca evidenziano la positiva incidenza che i sistemi di IA hanno avuto nella comprensione ed apprendimento della lingua, consentendo un miglioramento di capacità linguistiche e comunicative degli studenti (Zhang, Huang, 2024).

Nello studio di Jo, si osserva una disamina su come tali sistemi possano essere implementati nel processo educativo, in particolare egli analizza il contesto universitario, mettendo in evidenza gli elementi positivi che ha riscontrato: acquisizione e applicazione della conoscenza, in quanto tali sistemi permettono di assimilare nuove conoscenze e di applicarle nei contesti pratici; personalizzazione, l'IA ha la capacità di adattarsi ad esigenze specifiche, rendendo così l'apprendimento mirato, coinvolgente ed efficace; novità; poiché stimola curiosità ed interesse da parte dell'utilizzatore (Jo, 2024).

Da tali evidenze, si può affermare che i sistemi di intelligenza artificiale rappresentano uno strumento che, potenzialmente, potrebbe migliorare il processo di apprendimento e, conseguentemente, l'educazione digitale limiterebbe i possibili usi erronei della emergente tecnologia. Come sostiene Ciasullo *«i sistemi di IA sono potenti mezzi a fini educativi»* (Ciasullo, 2024: 72).

3. Il problema dell'antropomorfizzazione

Quel che oggi viene definito intelligenza artificiale dall'opinione pubblica sono i ChatBot. Questi sono sistemi con i quali tutti noi abbiamo facilità di interazione poiché vengono pubblicizzati sui social network e sono implementati direttamente nei motori di ricerca, caso emblematico è Copilot, nelle numerose applicazioni, come l'assistente IA di Acrobat Reader, o ancora nella maggior parte dei telefoni, basti pensare ad Apple Intelligence o

Gemini introdotte dai due colossi Apple e Google. Osservando queste evoluzioni e implementazioni, tutti noi siamo esposti all'intelligenza artificiale, in maniera diretta o indiretta, volontariamente e non.

I ChatBot sfruttano i modelli di grandi dimensioni, e vengono addestrati su una mole potenzialmente infinita di dati, e sono in grado di analizzare, in fase di input, e di elaborare, in fase di output, testi, audio, immagini. In aggiunta, una delle peculiarità di questi sistemi è il loro metodo di risposta, la quale è in grado di emulare, quasi perfettamente, risposte che sarebbero fornite da un essere umano. Questo processo, considerato dai pionieri della disciplina come fondamentale per definire una macchina "intelligente" (McCarthy *at al.*, 2006), ha prodotto dei cambiamenti significativi ed ha sollevato rilevanti perplessità sul fenomeno dell'antropomorfizzazione. Questo processo è antico quanto l'uomo poiché, l'uomo ha da sempre la tendenza a proiettare fattezze umane a varie entità, tra cui le divinità, un processo descritto a chiare lettere da Feuerbach nella Teogonia e in altri suoi scritti (Magris, 2020).

La caratteristica propriamente umana che spesso viene riconosciuta a questi modelli è l'intelligenza poiché riescono a trovare risposte, soluzioni a quesiti complessi ai quali, il solo utilizzatore, non riesce a fornire una risposta adeguata. Oltre a riconoscerne l'intelligenza, si tende a considerare questi modelli come sistemi relazionali, commettendo l'errore di interfacciarcisi come se lo fossero. (Brando, 2025). Questa problematica, come osservano Floridi e Nobre, è anche legata al *«conceptual borrowing»* che le nuove discipline adottano dai campi affini (Floridi, Nobre, 2024).

Il fenomeno in oggetto è stato elemento di interesse negli studi di Weizenbaum, nella seconda metà dell'900, che lo battezza 'Effetto ELIZA' (Natale, 2021), riprendendo il nome dal chatbot di assistenza psicologica che egli stesso ideò nel 1966. Come racconta anche nel suo libro "Computer power and human reason", Wiezenbaum rimase molto colpito dagli effetti del suo ChatBot perché, chi interagiva con esso, sembrava dimenticare che ELIZA fosse un programma informatico e tendeva ad attribuirle una vera e propria personalità (Wiezenbaum, 1976). L'evento descritto accadde alla sua segretaria, che iniziò a parlare con ELIZA e, nonostante fosse consapevole che si trattasse di un sistema informatico, si fece trasportare dalla discussione e chiese a Weizenbaum di lasciare la stanza (Weizenbaum, 1976). Come afferma egli stesso: «[…] *si sono lasciati coinvolgere emotivamente dal computer e lo hanno antropomorfizzato in modo inequivocabile*». (Weizenbaum, 1976: 6-7).

Spesso, anche ai nuovi chatbot viene attribuita, erroneamente, la capacità del pensiero autonomo, venendo scambiati per esseri umani, amici o

addirittura partner, con i quali poter avere un dialogo e con i quali intraprendere un rapporto intimo ed emotivo. Un drammatico esempio riguarda un quattordicenne statunitense che nel 2024 aveva instaurato una relazione intensa con un sistema di intelligenza artificiale, comunicando con esso quotidianamente. Questo rapporto virtuale era basato su un senso di empatia che il ragazzo percepiva, tanto da confidare al ChatBot i dettagli più intimi della sua vita privata. I suoi genitori erano all'oscuro di questa interazione quotidiana fino a quando il ragazzo non prese la drammatica decisione di togliersi la vita, influenzato dalla forte fiducia che aveva riposto nel sistema (Roose, 2024)

L'applicazione con la quale comunicava permette di personalizzare l'avatar con il quale si entra in contatto, facendogli imitare le fattezze di una persona che si vuole replicare, sia essa reale o di fantasia, di propria conoscenza o famosa, e permette di dialogarci, attraverso messaggi di testo o vocali, grazie al caricamento preventivo di files audio o testuali della persona che il sistema analizza e, riprendendone la cadenza e le parole maggiormente utilizzate, il sistema riesce a ricreare messaggi e una voce artificiale in grado di riprodurre la persona desiderata. L'esempio riportato è solo uno dei numerosi casi di un utilizzo errato dei sistemi e di una mancata attenzione verso l'etica di questi strumenti che oramai ci pervadono e sono a disposizione di ogni soggetto che naviga nel web, anche minorenne. Questi drammatici eventi pongono enfasi su problematiche sociali e giuridiche che devono essere poste all'attenzione di tutti noi, cercando, nel breve periodo, di trovare delle risoluzioni e di estendere l'AI Act per poter evitare che eventi di questo tipo possano diffondersi a macchia d'olio.

Conclusioni

I sistemi di intelligenza artificiale hanno portato a grandissimi progressi in svariati ambiti, dai settori lavorativi alla ricerca scientifica, ma, allo stesso tempo, hanno inevitabilmente generato nuove problematiche sociali di ampia portata, tra cui l'antropomorfizzazione, che può provocare danni sul lungo periodo, in quanto non permette di vedere la realtà dei sistemi e di quello che effettivamente sono e realizzano.

Più il processo di evoluzione di questi sistemi va avanti maggiore sarà la loro applicazione nell'ambito della vita quotidiana. L'obiettivo che occorre porsi e che si vuole cercare di delineare è come evitare che si possano ripetere eventi drammatici, come quello preso in esame in precedenza, cercando di indirizzare tali strumenti verso l'utilizzo etico e più incentrato sulla sfera

umana, l'elemento centrale che occorre preservare e attorno alla quale i sistemi di intelligenza artificiale devono gravitare.

Tutto questo può essere affrontato attraverso l'educazione al digitale, il cui utilizzo richiede una formazione adeguata, che deve partire da percorsi di alfabetizzazione e approfondimento relativi al funzionamento degli LLM (Ciasullo, 2024).

La sfida educativa, l'attenzione alla creazione di una consapevolezza di utilizzo di questi sistemi è la base per poter formare una generazione capace di destreggiarsi all'interno di una società sempre più complessa. La formazione permetterà di superare le diffidenze e comprendere i possibili risvolti (Ciasullo, 2024; Jo, 2024; Ait Baha *et al.*, 2024).

Come evidenziato da Floridi (2022), è essenziale riconoscere che l'intelligenza artificiale non rappresenta una forma autentica di intelligenza, ma piuttosto una modalità operativa, una riserva di capacità di agire che risulta efficace grazie all'adattamento dell'ecosistema umano e ambientale. Questa consapevolezza costituisce il fondamento per un utilizzo appropriato delle tecnologie digitali. Solo attraverso un approccio educativo integrato e strategico sarà possibile formare individui capaci di sfruttare il potenziale tecnologico in modo responsabile, contribuendo alla costruzione di una società digitale equa, consapevole e pronta ad affrontare le sfide future (Floridi, 2022).

Riferimenti bibliografici

Abdala Moreira K.A. (2023). Levels: The materiality of taste and artificial intelligence. *E|C*, (38): 286-295.

Adamopoulou E., Moussiades L. (2020). An overview of chatbot technology. In Maglogiannis I., Iliadis L., Pimenidis E. (a cura di), *Artificial Intelligence Applications and Innovations. AIAI 2020. IFIP Advances in Information and Communication Technology*, 584. Cham: Springer. https://doi.org/10.1007/978-3-030-49186-4_31

Ait Baha T., El Hajji M., Es-Saady Y. et al. (2024). The impact of educational chatbot on student learning experience. *Education and Information Technologies*, 29: 10153-10176. https://doi.org/10.1007/s10639-023-12166-w

Aragona B. (2020). Sistemi di decisione algoritmica e disuguaglianze sociali: le evidenze della ricerca, il ruolo della politica. *La Rivista delle Politiche Sociali*, 2(20): 213-226.

Bauman Z. (2011). *Modernità liquida*. Trad. di S. Minucci. Bari: Laterza.

Beck U. (2000). *La società del rischio: verso una seconda modernità*. Trad. di W. Privitera. Roma: Carocci.

Bennato D. (2024). *La società del XXI secolo*. Bari: Laterza.

Bennato D. (6 agosto 2024). Il sesso ai tempi del digitale: ecco le nuove forme di relazione. *Agenda Digitale*. https://www.agendadigitale.eu/cultura-digitale/il-sesso-ai-tempi-del-digitale-ecco-le-nuove-forme-di-relazione/

Biever C. (2023). ChatGPT broke the Turing test — the race is on for new ways to assess AI. *Nature*, 619(7971): 686-689. https://doi.org/10.1038/d41586-023-02361-7

Bostrom N. (2023). *Superintelligenza. Tendenze, pericoli, strategie*. Trad. di S. Frediani. Torino: Bollati Boringhieri.

Brando M. (2025). Amore, intimità e inganno nell'era dell'intelligenza artificiale. *Treccani Magazine*. https://www.treccani.it/magazine/atlante/societa/amore-intimita-e-inganno-nell-era-dell-intelligenza-artificiale.html

Castells M. (2014). *La nascita della società in rete*. Milano: Università Bocconi Editore.

Ciasullo A. (2024). Intelligenze artificiali in educazione: pensare oltre la fruizione. *RTH - Education & Philosophy*, 11. https://doi.org/10.6093/2284-0184/10755

Crawford K. (2021). *Né intelligente né artificiale. Il lato oscuro dell'IA*. Bologna: il Mulino.

Feuerbach L. (2010). *Teogonia: secondo le fonti dell'antichità classica, ebraica e cristiana*. A cura di A. Cardillo. Bari: Laterza.

Filimowicz M. (2022). *Deep Fakes: Algorithms and Society*. London: Routledge.

Floridi L., Nobre A.C. (2024). Anthropomorphising machines and computerising minds: The crosswiring of languages between artificial intelligence and brain & cognitive sciences. *Minds & Machines*, 34(5). https://doi.org/10.1007/s11023-024-09670-4

Floridi L. (2023). AI as agency without intelligence: On ChatGPT, large language models, and other generative models. *Philosophy & Technology*, 36: 15. https://doi.org/10.1007/s13347-023-00621-y

Floridi L. (2022). *Etica dell'intelligenza artificiale: sviluppi, opportunità, sfide*. Milano: Raffaello Cortina Editore.

Franklin D. (2022). *The Chatbot Revolution: ChatGPT. An In-Depth Exploration*. Independently Published.

Gobo G., Marcheselli V. (2021). *Sociologia della scienza e della tecnologia. Un'introduzione*. Roma: Carocci.

Intelligenza artificiale in Italia: mercato in crescita record. (22 ottobre 2024). *Osservatori Digital Innovation del Politecnico di Milano*. https://www.osservatori.net/comunicato/artificial-intelligence/intelligenza-artificiale-italia/

ITEA (2007). *Standards for Technological Literacy* (3ª ed.). International Technology Education Association.

Jo H. (2024). From concerns to benefits: A comprehensive study of ChatGPT usage in education. *International Journal of Educational Technology in Higher Education*, 21(1). https://doi.org/10.1186/s41239-024-00471-4

Kasneci E., Seßler K., Küchemann S., Bannert M., Dementieva D., Fischer F., Kasneci G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103: 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kissinger H.A., Schmidt E., Huttenocher D. (2024). *L'era dell'intelligenza artificiale*. Milano: Mondadori.

Kranzberg M. (1986). Technology and history: "Kranzberg's laws". *Technology and Culture*, 27(3): 544-560. https://doi.org/10.2307/3105385

Limna P., Kraiwanit T., Jangjarat K., Klayklung P., Chocksathaporn P. (2023). The use of ChatGPT in the digital era: Perspectives on chatbot implementation. *Journal of Applied Learning & Teaching*, 6(1). https://doi.org/10.37074/jalt.2023.6.1.32

Magris A. (2020). Il capolavoro dimenticato di Feuerbach: la *Teogonia*.

Microsoft News Center (2016). *Democratizing AI*. https://news.microsoft.com/features/democratizing-ai/

McCarthy J., Minsky M.L., Rochester N., Shannon C.E. (2006). A proposal for the Dartmouth Summer Research Project on Artificial Intelligence (August 31, 1955). *AI Magazine*, 27(4): 12. https://doi.org/10.1609/aimag.v27i4.1904

Natale S. (2021). The ELIZA effect: Joseph Weizenbaum and the emergence of chatbots. In *Joseph Weizenbaum and the Emergence of Chatbots*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780190080365.003.0004

Özçelik N.P., Ekşi G.Y. (2024). Cultivating writing skills: The role of ChatGPT as a learning assistant — a case study. *Smart Learning Environments*, 11(1). https://doi.org/10.1186/s40561-024-00296-8

Ranieri M. (2024). Intelligenza artificiale a scuola. Una lettura pedagogico-didattica delle sfide e delle opportunità. *Rivista di Scienze dell'Educazione*, LXII(1): 123-135.

Rocca G.G. (2024). IA e tecnologia: l'uso come attualizzazione di modelli culturali e sistemi semiotici. *E|C. Rivista dell'Associazione Italiana di Studi Semiotici*, 41: 1-18. Milano-Udine: Mimesis Edizioni. https://mimesisjournals.com/ojs/index.php/ec/article/download/4457/3473/

Roose K. (23 ottobre 2024). Can A.I. be blamed for a teen's suicide? *The New York Times*. https://www.nytimes.com/2024/10/23/technology/characterai-lawsuit-teen-suicide.html

Scarano R., Ferrantino C. (2024). Formare all'intelligenza artificiale: un progetto-studio con docenti e futuri docenti. *Education Sciences and Society*, 2: 72-87. https://doi.org/10.3280/ess2-2024oa18463

Shneiderman B. (2020). Human-centered artificial intelligence: Reliable, safe and trustworthy. *International Journal of Human-Computer Interaction*, 36(6): 495-504. https://doi.org/10.1080/10447318.2020.1741118

Stolpe K., Hallström J. (2024). Artificial intelligence literacy for technology education. *Computers and Education Open*, 6: 100159. https://doi.org/10.1016/j.caeo.2024.100159

Tarchi C., Zappoli A., Casado Ledesma L., Brante E.W. (2024). The use of ChatGPT in source-based writing tasks. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-024-00413-1

Triveri A., De Vivo M. (2024). Progettare le proteine: la rivoluzione computazionale. *La Chimica e l'Industria Online*, novembre-dicembre: 10-15.

van Dis E.A., Bollen J., Zuidema W., van Rooij R., Bockting C.L. (2023). ChatGPT: Five priorities for research. *Nature*, 614(7947): 224-226. https://doi.org/10.1038/d41586-023-00288-7

Weizenbaum J. (1976). *Computer Power and Human Reason: From Judgment to Calculation*. San Francisco: W.H. Freeman & Company.

Williams P.J. (2009). Technological literacy: A multiliteracies approach for democracy. *International Journal of Technology and Design Education*, 19: 237-254.

Zhang Z., Huang X. (2024). The impact of chatbots based on large language models on second language vocabulary acquisition. *Heliyon*, 10(3): e25370. https://doi.org/10.1016/j.heliyon.2024.e25370

L'Artificial Intelligence, il capitalismo delle piattaforme e i rischi per la democrazia

di *Sara Sbaragli*[*]

Il saggio analizza il rapporto tra intelligenza artificiale, capitalismo delle piattaforme e democrazia, mostrando come i sistemi algoritmici, lungi dall'essere neutri, riflettano interessi economici e politici. Richiamando gli sviluppi dell'AI e il ruolo delle Big Tech, si evidenzia come il capitalismo digitale produca diseguaglianze e processi pervasivi di datificazione. Il quadro teorico richiama Habermas e le prospettive critiche provenienti dalla sociologia digitale e dalla governamentalità algoritmica. L'articolo sottolinea infine i rischi per la deliberazione democratica, minacciata dalla sostituzione con sistemi predittivi e di nudging.

Parole chiave: intelligenza artificiale; capitalismo delle piattaforme; democrazia; algoritmi; sfera pubblica; Big Tech.

Artificial Intelligence, platform capitalism, and the risks to democracy

The article examines the relationship between artificial intelligence, platform capitalism, and democracy, showing how algorithmic systems, far from being neutral, embody economic and political interests. After outlining the developments of AI and the role of Big Tech, it highlights how digital capitalism generates inequality and pervasive processes of datafication. The theoretical framework draws on Habermas as well as critical perspectives from digital sociology and algorithmic governmentality. Finally, the paper stresses the risks for democratic deliberation, increasingly threatened by predictive and nudging systems.

Keywords: artificial intelligence; platform capitalism; democracy; algorithms; public sphere; Big Tech.

Introduzione

Gli interessi economici e politici hanno spostato le finalità dell'AI verso strumenti applicabili sul piano tecnologico. La crescita di investimenti e progetti in AI da parte del M.I.T., Stanford, I.B.M. e altre importanti uni-

DOI: 10.5281/zenodo.17524749

[*] Università di Napoli Federico II. sarasbaragli@gmail.com.

versità, istituti di ricerca, laboratori privati è esponenziale. Le grandi multinazionali come Google, Facebook, Amazon, Microsoft e Apple e altre emergenti sono colossali. Ogni ambito della riproduzione materiale e simbolica del mondo della vita è oggetto di *data analysis* di ingenti quantità di dati e *data-driven*, *customer service* tramite *chatbot* intelligenti e assistenti virtuali, *software* robotizzati e algoritmi di *machine learning,* protocolli di verifica predittiva. In questo scenario, l'AI non rappresenta soltanto un'evoluzione tecnologica, ma una trasformazione strutturale che incide sulle forme organizzative, produttive e culturali delle società contemporanee. Essa si configura come un'infrastruttura invisibile e pervasiva, capace di modificare i processi di conoscenza e di relazione, tanto nella sfera pubblica quanto nella vita quotidiana degli individui.

Il primo paragrafo offre una definizione provvisoria di intelligenza artificiale, delineando gli strumenti e le tecniche utilizzate nell'AI, compresi algoritmi di machine learning, reti neurali artificiali e altre metodologie avanzate che consentono alle macchine di apprendere e di adattarsi all'ambiente. L'AI sta già rivoluzionando numerosi aspetti della vita, migliorando l'efficienza e l'accuratezza in campi come la diagnosi medica, la gestione dei dati e l'automazione industriale. Questi progressi evidenziano il suo potenziale come strumento in grado di potenziare le capacità umane. Tuttavia, vengono sollevate anche preoccupazioni riguardo alla possibile disumanizzazione. Il secondo paragrafo analizzerà gli impatti sociali dell'AI sul lavoro, sull'economia e sulle relazioni quotidiane, con attenzione alla responsabilità decisionale e alla tutela di valori come dignità, autonomia e giustizia, garantiti dagli ordinamenti democratici. Saranno accennati gli impatti sociali dell'AI, in particolare sul mercato del lavoro, sulle dinamiche economiche e sulle interazioni umane quotidiane. Le implicazioni dell'utilizzo di sistemi autonomi sono oggetto di discussione, soprattutto in relazione alla responsabilità decisionale e alla conservazione di valori umani fondamentali come la dignità, l'autonomia e la giustizia – valori la cui realizzazione è compito degli ordinamenti politici democratici. Nell'ultimo paragrafo, infine, verrà discusso il quadro teorico di riferimento, che integra la riflessione sulla neutralità algoritmica con i contributi della sociologia digitale e della teoria della sfera pubblica, per mostrare come l'AI e il capitalismo delle piattaforme stiano ridisegnando i confini della democrazia contemporanea.

1. Gli sviluppi dell'intelligenza artificiale

Il termine *Artificial Intelligence* (AI) è in uso da molto tempo, ma ha suscitato l'interesse nell'ultimo decennio grazie al perfezionamento di tecnologie e applicativi. È una branca del Cognitive Computing (CC) che studia i processi mentali degli esseri umani al fine di migliorare l'interazione con le macchine. Questo tipo di tecnologia si basa su strumenti informatici dotati di grandi capacità di elaborazione di dati massivi (Big data) e di interazione con numeri elevati di variabili che consentono l'ottimizzazione dei processi di creazione di una molteplicità di risultati (output). Una prima classificazione di AI formulata dal filosofo John Rogers Searle (1990) che distingue tra Intelligenza artificiale "debole" (Weak AI) e Intelligenza artificiale "forte" (Strong AI).

Quest'ultima concerne dei sistemi esperti dotati di particolari capacità cognitive e che emulano in autonomia le migliori conoscenze e le abilità degli esseri umani. Si tratta ancora di un obiettivo ipotetico su cui lavorano accademici e imprese e che domina da tempo la fantascienza e la futurologia. Per contro, la Weak AI o Artificial Narrow Intelligence (ANI) sta avendo uno sviluppo applicativo esponenziale. Si basa sul "come se" vi fosse l'unione di pensiero e azione. È una forma di intelligenza programmata per svolgere determinate attività complesse per l'uomo, quali ad esempio la traduzione di testi o la risoluzione di calcoli. Non si tratta di software dotati di intelligenza umana ma di sistemi in grado di assistere l'essere umano, eseguendo in modo rapido e preciso svariate operazioni ed emulando il comportamento umano. Ad esempio, i sistemi Weak AI sono presenti nel filtro delle email, nei servizi di raccomandazione (come Spotify), negli assistenti vocali (come Siri e Alexa per riconoscere la voce dell'utente, elaborare le richieste e fornire le risposte appropriate), in campo medico, per esempio per l'analisi di immagini di risonanza magnetica o di tomografia computerizzata, nella produzione di automobili, nella gestione di magazzini e in molti altri.

Il funzionamento dipende da quattro funzioni specifiche tra loro strettamente interconnesse: ascolto, comprensione, apprendimento ed interazione. Ascolto riguarda la capacità della macchina di acquisire le informazioni circostanti e raccogliere e classificare i dati organizzandoli nel miglior modo possibile. La comprensione è relativa alla capacità di analisi ed elaborazione dei dati acquisiti nella fase di ascolto, elaborando ripetizioni di schemi e fornendo le informazioni necessarie all'essere umano per poter prendere decisioni efficaci. Tali dati non sono considerati singolarmente e catalogati, ma raccolti e messi in relazione tra loro al fine di migliorare significativa-

mente l'intervento. L'apprendimento concerne la capacità della macchina di apprendere nella misura in cui riesce a imparare a svolgere determinate funzioni e compiti attraverso l'analisi e l'elaborazione dei dati precedentemente raccolti (*Machine Learning*). Dall'evoluzione di tale apprendimento si è sviluppato il concetto di Deep Learning o "apprendimento profondo", associato alle "reti neurali". L'interazione, infine, si ha se la macchina è in grado di prendere una decisione e interagire con l'uomo. La tecnologia che rende possibile tutto ciò è il *Natural Language Processing* (NLP). Un'altra applicazione di *Deep Learning* è la *Computer Vision* (CV) che consente alcune funzionalità come il rilevamento di informazioni sensoriali, il riconoscimento dell'immagine termica e il riconoscimento facciale.

Questa forma di AI è entrata nel funzionamento dei sistemi economici e amministrativi e nella riproduzione delle nostre forme di vita sia nella sfera pubblica che in quella privata. Le imprese stanno reinventando i processi produttivi, organizzativi e commerciali e i settori in cui applicano le tecnologie offerte da questa rivoluzione dirompente sono ormai molteplici. L'intelligenza artificiale si è così innestata nel tessuto dei sistemi organizzativi e delle forme di vita quotidiana da divenire una realtà immanente, invisibile e inconsapevole, una sorta di inconscio tecnologico. E tuttavia la programmazione dei sistemi di AI che costruisce questo mondo ibrido non è "neutrale": «rispecchia e ricalca i principi valoriali e comportamentali di uno specifico gruppo sociale e culturale: la classe dei soggetti leader nel campo hi-tech che ha dato vita ad un sistema algoritmico capitalista» (Grassi, 2024: 10). Il mito della neutralità algoritmica (Airoldi, Gambetta, 2018) è alla base dell'ideologia del riduzionismo scientista sodale con il capitalismo delle piattaforme. Il capitalismo delle piattaforme mostra che l'AI non è solo progresso tecnico, ma un processo storico-sociale che riflette rapporti di potere e valori dominanti. Diventa un dispositivo culturale e politico, un "inconscio tecnologico" che orienta pratiche e decisioni, aprendo opportunità ma anche rischi per democrazia e autonomia.

2. Il capitalismo delle piattaforme

Oggi si assiste a una concentrazione di potere nelle grandi corporazioni digitali e nelle reti. Con il concetto di Big Tech ci si riferisce alle grandi aziende tecnologiche multinazionali – Alphabet (Google, YouTube), Amazon, Apple, Meta (Facebook, Instagram, WhatsApp e Messenger) e Microsoft. Un gruppo più ampio, chiamato i Magnifici Sette, aggiunge alle GAFAM anche Nvidia e Tesla. Vi è una parziale coincidenza con i produt-

tori di AI, con l'aggiunta di Databricks, Anduril Industries, Gong, Anthropic, IBM, Huawei. Sarebbe un'illusione pensare che le piattaforme regolate da algoritmi siano prive di interessi e forme di controllo senza effetti sulla programmazione. Al contrario, essi sono i custodi della nostra forma di vita. Per Srnicek, le piattaforme mostrano una spinta colonizzatrice che non le vede «accontentarsi di dominare il mercato, ma puntano a diventare il mercato stesso» (2017: 47). Non sono operatori di un mercato in cui si esplica la libera attività dell'*homo oeconomicus* ma un'infrastruttura il cui assetto valoriale e operativo è definito dalle piattaforme (Corchia, Borghini, 2025). Hanno saturato l'intero orizzonte socio-tecnico in cui si muovono attori istituzionali, imprese e utenti, svolgendo funzioni di intermediazione centrale e inedita che struttura il flusso informativo tramite la logica degli algoritmi (quasi sempre inavvertita e non trasparente agli utenti). Come scrive Colin Crouch, quella che «sembrava essere una tecnologia di liberazione e democrazia finisce così per favorire un manipolo di individui e gruppi estremamente ricchi» (2020/2020: 6-7). Le applicazioni dei sistemi di AI stanno generando o accentuando degli squilibri nella dotazione delle chance di vita, patologie nella riproduzione materiale e simbolica dei sistemi sociali e dei mondi vitali e nel rapporto ecologico tra specie umana e ambiente naturale.

Consideriamo solo gli effetti appariscenti. Una prima questione riguarda la perdita di posti di lavoro, la crisi occupazionale causata dal saldo negativo prevedibile tra nuovi e vecchi lavori investiti dall'AI. L'accesso alla ricchezza da lavoro, infatti, è stato la leva con cui gli Stati hanno assicurato un benessere materiale diffuso nella popolazione e realizzato il compromesso sociale pur non intaccando il solido nocciolo classista. Il crescere della disoccupazione tecnologica che tocca sia le professioni manuali che quelle intellettuali. Un altro macro-fenomeno concerne le diseguaglianze nei modi di produzione tra le imprese e tra i sistemi paese con e senza questa nuova tecnologia. Infatti, è necessario disporre di un'infrastruttura informatica (*hardware* e *software*) di calcolo per eseguire le applicazioni di intelligenza artificiale e addestrare i modelli di machine learning. Per addestrare sistemi di AI è necessario un enorme volume di dati. Si deve disporre di una capacità di archiviazione per gestire ed elaborare i dati. Poiché i sistemi di AI possono creare degli output apparentemente autorevoli come constatazioni o affermazioni, è importante inoltre verificarne la fonte e la veridicità. I costi per l'acquisto delle macchine e dei programmi, l'impegno per la riorganizzazione aziendale, la gestione dei processi, la formazione del personale e i servizi esterni specializzati possono essere molto elevati e diminuire la redditività dell'investimento in AI per imprese e amministrazioni. Ciò accresce la dipendenza dai servizi disponibili delle grandi multinazionali del settore. Infi-

ne, emerge l'inarrestabile erosione della sovranità nazionale non solo nei Paesi tecnologicamente dipendenti dalle potenze informatiche. L'influenza delle corporazioni high-tech ha spostato il potere conoscitivo-comunicativo e il centro di controllo dalla dimensione statuale, con i propri sistemi di *check and balance*, a quella dell'economia sovranazionale. È il dominio che Shoshana Zuboff definisce "capitalismo della sorveglianza" (2019/2023).

Se il dibattito sulle piattaforme e sull'AI si è concentrato principalmente su scala macro (ovvero sulla loro economia politica, sulle loro conseguenze più ampie sulla circolazione dei contenuti, sul loro impatto sui valori condivisi), occorre svolgere un ampio lavoro teorico ed empirico per comprendere in che modo le esperienze e le pratiche concrete delle persone vengono effettivamente rimodellate e co-evolvono (Magaudda, Piccioni, 2019). Massimo Airoldi descrive questa profilazione con il termine *machine habitus*: «prodotto e adattato sulla base di dati socialmente strutturati generati dagli utenti, contribuisce a sua volta a produrre e regolare le pratiche e le disposizioni degli utenti, esercitando un'autorità computazionale opaca» (2024: 165). Ai modelli di datificazione sempre più complessi si accompagna l'uso di algoritmi senza che vi sia consapevolezza di quanto la vita quotidiana sia piena di contenuti predefiniti. La disponibilità di big data è un prerequisito. Ciò viene assicurato grazie ai dati donati, a nostra insaputa o con il nostro consenso in un processo che, dal punto di vista delle piattaforme, Marion Fourcade e Daniel N. Kluttz definiscono "accumulazione per dono" (2020). I *recommendation system* delle piattaforme possono influenzare gli utenti invadendo persino la loro autonomia decisionale: indirizzandoli verso determinati contenuti o limitandone la varietà. Uno dei maggiori fenomeni legati a questo concetto è il *nudging*, una "spinta gentile", che consiste nell'applicazione di rinforzi positivi o negativi per incidere sulle preferenze e i processi decisionali dell'individuo o di gruppo. Nell'ambito dell'intelligenza artificiale, il fenomeno si converte in *hypernudging* e presenta due obiettivi: semplificare il processo decisionale dell'utente e prevedere le sue azioni. Ciò giustifica la ripresa di concetti e prospettive foucaltiane per comprendere la trama infrastrutturale del capitalismo contemporaneo rispetto al nesso tra i modi di produzione del sapere e i modi di esercizio del potere. Con la governamentalità algoritmica il mercato estende progressivamente la sua logica di funzionamento e la propria "arte di governo" a tutte le altre sfere della vita sociale in un processo di "mercantizzazione del mondo" (Chicchi, 2024).

Il capitalismo delle piattaforme non riguarda solo l'economia, ma penetra nelle pratiche quotidiane, orientando scelte e comportamenti tramite profilazione e previsione. L'illusione della neutralità tecnologica lascia

spazio a un'infrastruttura che concentra potere e limita l'autonomia. La sociologia è così chiamata a riflettere sia sugli assetti macro del digitale sia sulle nuove forme di soggettivazione che esso genera.

3. Un attacco al nucleo deliberativo della democrazia

Nel senso ampio di infrastrutture del pensiero e dell'azione, aperture epistemiche e modi d'uso pragmatici, i sistemi di AI sono degli "algoagenti" che ridefiniscono i confini di pensabilità e realizzabilità delle esperienze in un mondo ampiamente artefatto. In tale accezione, per Keller Easterling, un'infrastruttura genera una disposizione nella misura in cui è un «modo di organizzazione» che consiste nel «fare attivamente qualcosa» (2014: 73). Vando Borghi conferma che «Oltre alla valenza per cui più convenzionalmente sono tematizzate, cioè quella strumentale, in quanto sistemi sociotecnici di cui ci serviamo per fare molte cose, esse vanno infatti considerate anche come dispositivi che fanno essi stessi qualcosa di noi, che modificano in profondità le nostre forme di vita e le nostre pratiche sociali (2021).

L'effetto costituente delle infrastrutture è evidente se consideriamo il rapporto con il capitalismo sia come modo di produzione che come forma fenomenica di riproduzione dell'esistenza dei lavoratori e dei consumatori. Il capitalismo delle piattaforme, con le regolazioni e previsioni algoritmiche, mina le condizioni minime di riproduzione del mondo della vita generando effetti patologici in tutte le dimensioni culturali, sociali ed esistenziali: perdita di senso, insicurezza nell'identità collettiva, anomia nei legami di appartenenza, crisi di orientamento e di educazione, disturbi psicologici, etc. Vi sono effetti evidenti anche sul piano pubblico dei processi democratici. Sostituire la volizione e la scelta individuale non riguarda solo il consumo di prodotti culturali, film, musica, etc. ma si espande al cuore della democrazia. Dentro la cornice formale dei princìpi liberali di rappresentanza, divisione dei poteri e tutela delle minoranze, la sostanza repubblicana della formazione della volontà e delle decisioni viene prosciugata del momento deliberativo.

Sin dagli anni Ottanta, si assiste a un cambiamento del sistema dei media lungo le direttrici della decentralizzazione, diversificazione e personalizzazione sospinte dai nuovi media che cominciano a mutare l'esperienza audiovisiva del pubblico di massa. L'offerta si è pluralizzata e diversificata. Oltre all'overdose da esposizioni, la crescita del sistema mediale si è accompagnata alla differenziazione della sfera pubblica, con molteplici media, fonti, forme e contenuti – e alla "frammentazione" del pubblico media-

le. Il mega-testo della sfera pubblica generale è stato incessantemente riscritto in un numero indefinito di micro-testi e in una miriade di circuiti che tematizzano questioni sovente intrecciate di cronaca, politica, economia, cultura, sport o intrattenimento e che, attraverso ponti ermeneutici, definiscono sfere pubbliche parziali, porose, più o meno specializzate e in collegamento tra loro. I sistemi di intelligenza artificiale accentuano esponenzialmente la frammentazione della sfera pubblica e la creazione algoritmica di *filter bubble* in un processo centrifugo che porta a isolarsi nel perimetro iper-personalizzato delle selezioni condizionate da logiche di filtraggio invisibili (Pariser, 2011). I sistemi di filtraggio isolano gli utenti, limitando la loro esposizione a contenuti e prospettive, determinando così effetti dannosi nella formazione della volontà e dell'opinione pubblica e per le istituzioni democratiche (Fuchs, 2023).

Come osserva Kopsaj, lo spazio comunicativo delle piattaforme può essere efficacemente descritto mediante la metafora della "casa sull'albero": uno spazio elevato che offre nuove possibilità di visione e di connessione, ma che rimane al tempo stesso selettivo, segnato da accessi privilegiati ed esclusioni strutturali. L'immagine restituisce bene l'ambivalenza delle piattaforme e dei sistemi di intelligenza artificiale: promettono inclusione e apertura universale, ma in realtà istituiscono barriere invisibili, riproducendo e amplificando diseguaglianze già esistenti (2025). Viene meno l'idea di sfera pubblica che Habermas definiva «il sostrato organizzativo di un universale "pubblico di cittadini" emergente, per così dire, fuori dalla sfera privata» (1992/1996: 435). Uno spazio di intermediazione in cui circolano informazioni, si svolgono confronti, generano e aggregano modelli interpretativi, credenze generali, descrizioni e rappresentazioni di fatti, concezioni morali, valutazioni etiche, espressioni emotive e altre forme più o meno discorsive sui temi – attori, oggetti, eventi – più o meno controversi. Alla sfera pubblica sono affidate le funzioni di: 1) "filtro", in quanto seleziona solo alcune informazioni e opinioni dal "flusso babilonico di voci" che circolano nella società e sulle questioni di interesse più generale; 2) un "condensatore", in quanto presenta queste informazioni e opinioni sotto forma di possibili discussioni, traducendo il linguaggio comune in argomenti concorrenti; 3) una "cassa di risonanza", nel senso che raccoglie e dà visibilità alle informazioni e opinioni diffuse nella società, favorendo una comunicazione di tipo *botton up*; 4) "sistema di allarme", nel senso che mette le istituzioni pubbliche – chi prende decisioni – di fronte alle questioni più urgenti per i cittadini e gli interessi organizzati; 5) "sistema di controllo", che attiva i circuiti della *accountability* e *responsibility*, limitando l'autoreferenzialità delle classi dominanti e la corruzione dell'interesse

generale. La sfera pubblica è il "dominio del discorso", dove gli attori sociali imparano attraverso il dibattito tra argomenti diversi a confrontare e modificare i propri punti di vista, in cui si tematizzano, oltre alle pretese di verità e efficacia delle azioni tecniche, le pretese di bontà, giustezza, preferibilità, gusto e autenticità delle scelte. Essa è quindi un "generatore di apprendimenti" che accresce il capitale culturale e un "*medium* di coesione", in quanto l'estensione della densità delle relazioni di comprensione reciproca e intesa comunicativa favorisce la crescita di "capitale sociale" e in una certa misura vincoli solidaristici (Corchia, Bracciale, 2020). L'autochiarificazione genera conoscenze di sé e del mondo capaci di orientare l'agire nella relazione tra i fini e i valori e non solo in relazione tra i mezzi e i fini per il cui supporto serve l'*expertise* artificiale. La sfera pubblica diviene lo spazio in cui si coltivano le conoscenze e abilità critiche volte alla chiarificazione dei rapporti di dominio materiali e simbolici che ostacolano l'emancipazione degli esseri umani. La sfera pubblica è proprio questo spazio in cui avviene anche il disvelamento dei contenuti ideologici che sottraggono alla tematizzazione gli ordinamenti sociali. E dove si possono progettare "alternative all'esistente", ampliare i margini della visione del mondo diffondendo così la ricerca di senso per "ciò che manca" e "ciò che potrebbe essere altrimenti" (Habermas, 2006/2011). Le scelte valoriali sull'"essere-sé stessi", cosa possiamo e vogliamo essere, richiedono prese di coscienza individuali e collettive e – avvertiva Ardigò – non potranno mai essere compiute per noi da «"doppi" computerizzati» (1983). L'ideale di un pubblico che si risparmia lo sforzo di scegliere la propria forma di vita delegando agli algoritmi predittivi la volontà generale è disumano e funzionale agli interessi del nuovo capitalismo post-liberale. Di fronte al progetto di sostituzione del potere politico con un ordine tecnocratico si dovrebbe prendere posizione seguendo l'esempio di Luciano Floridi: «Con il rapido ritmo del cambiamento tecnologico, si è tentati di ritenere il processo politico nelle odierne democrazie liberali antiquato, sorpassato e non più all'altezza del compito di preservare i valori e di promuovere gli interessi della società e dei suoi membri. Non sono d'accordo» (2022/2022: 250).

Riferimenti bibliografici

Airoldi M. (2024). *Machine Habitus. Sociologia degli algoritmi*. Milano: Luiss University Press.

Airoldi M., Gambetta D. (2018). Sul mito della neutralità algoritmica. *The Lab's Quarterly*, 20(4): 25-46.

Ardigò A. (1983). Un nuovo processo mimetico: le ricerche di «intelligenze artificiali»:

interrogativi ed ipotesi di rilevanza. *Studi di Sociologia*, 21(3): 233-244.

Borghi V. (2021). Capitalismo delle infrastrutture e connettività. Proposte per una sociologia critica del mondo a domicilio. *Rassegna Italiana di Sociologia*, 3: 671-999.

Chicchi F. (2024). L'infrastruttura della società automatica: governamentalità algoritmica e capitalismo digitale. In Borghi V., Leonardi F. (a cura di), *Il sociale messo in forma. Le infrastrutture come cose, processi e logiche della vita collettiva* (pp. 149-166). Napoli-Salerno: Orthotes.

Corchia L., Borghini A. (2025). Infrastructure as a sociological category: concept, applications, and paradigmatic turns? *Journal of Classical Sociology*, 25(2): 123-151.

Corchia L., Bracciale R. (2020). La sfera pubblica e i mass media. Una ricostruzione del modello habermasiano. *Quaderni di Teoria Sociale*, 20(1-2): 353-381.

Crouch C. (2020). *Combattere la postdemocrazia*. Roma-Bari: Laterza.

Easterling K. (2014). *Extrastatecraft: The Power of Infrastructure Space*. London-New York: Verso Books.

Floridi L. (2022). *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*. Milano: Raffaello Cortina Editore.

Fourcade M., Kluttz D.N. (2020). A Maussian bargain: accumulation by gift in the digital economy. *Big Data & Society*, 7(1): 1-16.

Fuchs C. (2023). *Digital Ethics. Media, Communication and Society. Volume Five*. London-New York: Routledge.

Grassi E. (2024). *Sociologia algomorfica. Il ruolo degli algoritmi nei mutamenti sociali*. Milano: FrancoAngeli.

Habermas J. (1992). *Fatti e norme. Contributi a una teoria discorsiva del diritto e della democrazia*. Milano: Guerini e Associati.

Habermas J. (2006). La democrazia ha anche una dimensione epistemica? Ricerca empirica e teoria normativa. In Id., *Il ruolo dell'intellettuale e la causa dell'Europa* (pp. 63-108). Roma-Bari: Laterza, 2011.

Kopsaj V. (2025). *Digital age and inclusive future. Society, self and health*. Milano: FrancoAngeli.

Magaudda P., Piccioni T. (2019). Practice theory and media infrastructures. *Sociologica*, 13(3): 45-58.

Pariser E. (2011). *The Filter Bubble: What the Internet is Hiding from You*. London: Penguin.

Searle J.R. (1990). Is the brain's mind a computer program? *Proceedings and Addresses of the American Philosophical Association*, 64(3): 21-37.

Srnicek N. (2017). *Platform Capitalism*. Cambridge: Polity Press.

Zuboff S. (2019). *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*. Milano: LUISS University Press.