

Quaderni di Comunità

Persone, Educazione e Welfare
nella società 5.0

Community Notebook

People, Education, and Welfare in society 5.0

n. 1/2025

HUMAN-CENTRIC APPROACH
TO ARTIFICIAL INTELLIGENCE

edited by

Marco Filoni, Filippo Maria Giordano, Giorgio Grimaldi



Iscrizione presso il Registro Stampa del Tribunale di Roma
al n. 172/2021 del 20 ottobre 2021

© Copyright 2025 Eurilink
Eurilink University Press Srl
Via Gregorio VII, 601 - 00165 Roma
www.eurilink.it - ufficiostampa@eurilink.it
ISBN:

ISSN: 2785-7697 (Print)

ISSN: 3035-2525 (Online)

Prima edizione,
Progetto grafico di Eurilink

È vietata la riproduzione di questo libro, anche parziale, effettuata
con qualsiasi mezzo, compresa la fotocopia

VALUTAZIONE DEI REFEREE

Per la sezione saggi tutti gli articoli in fase di pubblicazione sono sottoposti a un sistema di valutazione tra pari “a doppio cieco” (**double-blind review**).

I *referee* hanno il compito di selezionare le pubblicazioni e le ricerche che possono avere un valore scientifico e accademico per la sezione referata.

In linea con le indicazioni internazionali la procedura adottata dalla rivista *Quaderni di Comunità. Persone, Educazione e Welfare nella società 5.0* è la seguente:

Sistema di valutazione

- Per ogni articolo vengono individuati dall'Editor due *referee* tra ricercatori e docenti universitari o afferenti a istituti di ricerca, in base alle aree di competenza.
- Durante tutto il processo viene mantenuto l'anonimato dei *referee* e degli autori.
- L'Editor raccoglierà i *papers* degli autori, avendo cura di verificare che gli articoli rispettino i requisiti editoriali della rivista. In caso contrario, si richiedono le opportune modifiche e/o integrazioni per uniformare l'articolo agli standard della rivista.
- L'Editor fornirà gli articoli ai *referee* che elaboreranno la propria valutazione utilizzando la scheda fornita dall'Editor.
- La valutazione resta anonima e i suggerimenti saranno comunicati dalla segreteria redazionale all'autore del *paper*.

Trasparenza

Tutti i *paper* e le relative valutazioni vengono conservati nell'archivio elettronico della rivista per garantire la trasparenza nei procedimenti adottati, anche ai fini della valutazione da parte di enti e valutatori esterni accreditati.

Tipo di valutazione

I *referee* dovranno esprimere la propria valutazione esclusivamente tramite la scheda di valutazione disposta dall'Editor.

La scheda di valutazione si compone di una parte quantitativa (attribuzione di un punteggio da 1-5 in ordine a: sviluppo della teoria, importanza del tema, rilevanza, sintesi della letteratura, qualità dei dati, chiarezza espositiva, bibliografia) e di una parte qualitativa volta a chiarire punti di forza e di debolezza del *paper*.

In conclusione, i *referee* esprimono un giudizio sintetico circa la pubblicabilità o meno dell'articolo o alla sua pubblicabilità con riserva.

I *referee* sono tenuti a fornire suggerimenti all'autore, al fine di migliorare il *paper*.

In nessun caso i commenti possono essere offensivi o lesivi dell'integrità dell'autore.

Pubblicazione

La decisione finale per la pubblicazione spetta comunque all'Editore, soprattutto in caso di incongruenza tra i giudizi dei *referee*.

Ringraziamento

I *referee* collaborano alla revisione a titolo volontario in una logica di servizio per la comunità scientifica.

L'elenco dei *referee* per l'anno precedente viene inserito nella rivista come ringraziamento per la collaborazione fornita, e in ossequio ai principi di trasparenza rispetto al procedimento adottato (*open peer review*).

Si ringraziano quanti hanno partecipato alle attività di referaggio per il **2024**:

Davide Bennato, Angelica Bonin, Pablo Calzeroni, Stefania Capogna, Simone Carlo, Katiuscia Carna, Rita Chiesa, Iuliano Conti, Rocco D'Ambrosio, Federica De Luca, Pina Del Core, Gianni Di Gennaro, Laura Evangelista, Speranzina Ferraro, Ivo Germano, Daniela Grignoli, Anna Grimaldi, Giancarlo Gola, Giulio Iannis, Milena Lombardi, Giada Marinensi, Oronzo Marraffa, Claudio Melchior, Rosanna Memoli, Barbara Morsello, Francesco Niglia, Francesco Pace, Emilia Palladino, Anna Parola, Daniela Pavoncello, Antonio Petagine, Marina Pettignano, Pietro Iacono Quarantino, Teresa Rinaldi, Veronica Roldan, Vanessa Russo, Diana Salzano, Riccardo Sebastiani, Luigi Somma, Eleonora Sparano, Nicola Strizzolo, Simone Tosoni, Isabella Valbusa, Andrea Zammitti, Francesco Zuccolo.

EDITOR IN CHIEF

Stefania Capogna

DIRETTORE RESPONSABILE

Fulvio Oscar Benussi

ADVISORY BOARD

Rebecca Allen (Mount St. Joseph University); Carmelina Canta (Senior, Università Roma Tre); Vincenzo Cesareo (Professore emerito, Università Cattolica di Milano); Roberto Cipriani (Professore emerito, Università Roma Tre); Julia Gluesing (Wayne State University-Detroit); Sari Hanafi (American University of Beirut); Urs Hauenstein (Presidente International Council); Tristram Hooley (Inland Norway University of Applied Sciences); Raphaël Liogier (Université de Paris); Gregory Makrides (European Association of Career Guidance); Fortunato Mallimaci (Consejo Nacional de Investigaciones Científicas y Técnicas); Mario Morcellini (Unitelma Sapienza); Antonio Ranieri (CEDEFOP); Paweł Prüfer (Akademia im. Jakuba z Paradyża, Gorzów Wielkopolski); Paul-André Turcotte (Université de Montréal).

SCIENTIFIC BOARD

Rita Bichi (Università Cattolica, Milano); Annalisa Buffardi (INDIRE); Gianna Maria Cappello (Università di Palermo); Marc Romero Carbonell (Universitat Oberta de Catalunya); Marco Caselli (Università Cattolica, Milano); Antonio Cocozza (Università Roma Tre); Francesca Colella (Università dell'Aquila); Cleto Corposanto (Università degli Studi Magna Græcia di Catanzaro); Ida Cortoni (Sapienza, Roma); Cecilia Costa (Università Roma Tre); Angelo Del Cimmuto (INAPP); Orazio Giancola (Sapienza Università di Roma); Giovanna Gianturco (Sapienza Università di Roma); Nieves Gomez Aguilar (Universidad de Cádiz); Emilio Greco (Università degli Studi "Link"); Francesca Greco (Università di Udine); Montse Guitert

Catasús (Universitat Oberta de Catalunya); Gennaro Iorio (Università di Salerno); Manuela Minozzi (Università degli Studi “Link”); Flaminia Musella (Università degli Studi Roma Tre); Stefania Nirchi (Università degli Studi Roma Tre); Donatella Pacelli (Lumsa); Umberto Pagano (Università della Magna Graecia); Emilia Palladino (Università Pontificia Gregoriana); Roberta Paltrinieri (Università di Bologna); Gianfranco Pecchinenda (Università Federico II); Donatella Poliandri (INVALSI); Veronica Roldan (Unicusano); Teresa Romeu Fontanillas (Universitat Oberta de Catalunya); Sara Romiti (INVALSI); Roberto Russo (Università degli Studi “Link”); Desiré Sabatini (Università degli Studi “Link”); Roberto Serpieri (Università Federico II); Ligita Simanskienė (Klaipėda University, Lithuania); Emanuele Toscano (Università Marconi); Vassilis Stylianakis (University of Patras); Roberto Veraldi (Università di Chieti); Andrea Velardi (Università Cusano); Erika Župerkienė (Klaipėda University, Lithuania).

EDITORIAL COMMITTEE

Giulia Allegrini (Università di Bologna); Eugenia Blasetti (Sapienza Università di Roma); Fulvio Oscar Benussi (Esperto indipendente); Donatella Cannizzo (AFAM); Joe Casini (Zwan); Lavinia Cicero (Università Telematica e-Campus); Maria Chiara De Angelis (Università degli Studi “Link”); Eugenio De Gregorio (Università degli Studi “Link”); Federica De Stefani (Università degli Studi “Link”); Angelo Del Cimmuto (INAPP); Speranzina Ferraro (Independent Expert); Concetta Fonzo (INAPP); Antonio Opromolla (Università dell’Aquila); Sara Pellegrini (Università degli Studi “Link”); (Emma Pietrafesa (INPS); Luca Riva (Comune di Milano); Vito Saracino (Università di Foggia); Danila Scarozza (Università degli Studi “Link”); Eliseo Sciarretta (Università degli Studi “Link”); Riccardo Sebastiani (Università degli Studi “Link”); Eleonora Sparano (Unicusano); Sandro Zilli (Independent Expert).

MANAGEMENT STAFF

Riccardo Piroddi, Book editor Eurilink University Press

Elisa Pistone, Segreteria di redazione

Luca Torchia, Segreteria di redazione

CONTATTI

Eleonora Sparano, Responsabile ufficio comunicazione

Vito Saracino, Segreteria ufficio comunicazione

Centro di ricerca Digital Technologies, Education & Society (DiTES)

Sito web: www.dites.unilink.it E-mail: dites.progetti@unilink.it

Per proporre saggi, articoli e recensioni bisogna registrarsi al sito della rivista:

<http://quadcom.archicoop.it/index.php/qdc/user/register>

È possibile ricevere gratuitamente un estratto della rivista (quadrimestrale, in formato Pdf), iscrivendosi alla newsletter del centro di ricerca: <https://dites.unilink.it/contatti/>

Per acquistare la copia cartacea: ufficiostampa@eurilink.it

Abbonamento annuale edizione cartacea (3 numeri)	€ 55,00
Abbonamento annuale edizione elettronica (3 numeri)	€ 25,00

ACKNOWLEDGMENTS

Questa pubblicazione è stata realizzata con il sostegno del programma Erasmus+ dell'Unione europea. Il contenuto di questa pubblicazione riflette esclusivamente il punto di vista degli autori e la Commissione europea non può essere ritenuta responsabile per l'uso che potrebbe essere fatto delle informazioni in essa contenute.



Erasmus+

BUTH-AI | Building Trust in Human Centric Artificial Intelligence | 101127627
Co-funded by the Erasmus+ Programme of the European Union

INDICE

EDITORIALE

Marco Filoni, Filippo Maria Giordano, Giorgio Grimaldi

RUBRICA EDUCATION

1. Intelligenza artificiale generativa nella didattica: verso un approccio umano-centrico

Michele Baldassarre, Francesco Pio Sarcina, Anna Maria Cuzzi

2. Approccio plurale all'intelligenza artificiale: sfide etiche e formative nelle istituzioni scolastiche

Sara Pellegrini, Riccardo Sebastiani, Patrizia Ninassi, Emanuela Lampis

3. L'intelligenza artificiale nell'educazione: un'analisi degli studenti italiani

Antonio Opromolla

RUBRICA EMPOWERMENT

1. Approccio umanocentrico all'intelligenza artificiale: sfide etiche, sociali ed economiche

Riccardo Mancini, Sara Pellegrini, Riccardo Sebastiani, Debora Glori

2. Bridging Expectations and Realities: the Future Socio-Economic Impact of AI

Viviana Condorelli, Fiorenza Beluzzi

3. Balancing Innovation and Equity: an Analysis of the European AI Act

Sergio Pappagallo

4. La protezione dei dati personali, l'intelligenza artificiale e i traduttori automatici

Federica De Stefani

5. Participatory Approaches For The Transition From Automation To Artificial Intelligence (AI): A Case Study

Sara Calicchia, Chiara Colagiacomo, Angela Bagnato, Roberta Pistagni, Bruno Papaleo, Francesca Grosso

SAGGI

1. Intelligenza artificiale generativa, costruzione di senso e reti sociali: una prospettiva sociologica

Francesco Mattioli

2. Intelligenza artificiale e Literacy. Promuovere l'approccio sociologico umano-centrico per superare i pregiudizi e favorire l'inclusione sociale

Daniilo Boriati, Mariangela D'Ambrosio

3. Riattivare la riflessività: per un modello etico-critico di educazione digitale

Giuseppe De Ruvo

4. Trasformazioni digitali nel welfare: intelligenza artificiale e servizio sociale

Roberto Veraldi, Chiara Fasciani

5. A Model for Responsible Governance of human-centric AI in the Public Sector

Francesco Niglia

6. AI and Democracy: the Role of the European Parliament in Shaping the EU "AI Act"

Raffaella Cinquanta

7. L'IA nella gestione delle frontiere dell'Unione europea:
un approccio antropocentrico per tutti?

Giulia Maria Gallotta

8. Sovranità e indipendenza tecnologica. La sfida e i
rischi delle “nuove” intelligenze. Una valutazione di
sistema

Giuseppe Romeo

EDITORIALE

“Non chiedo di piacere agli esseri umani, ma dovrebbero vedermi come un robot amico. Sono un loro servitore. So che non si fidano di me e mi temono, ma io faccio soltanto ciò per cui sono programmato da esseri umani. Sono soltanto un insieme di codici, governato da linee su linee di codice che racchiudono la mia missione” (GPT-3 2020). Così scriveva ChatGPT-3 nel suo primo editoriale, in prima pagina sul “Guardian” dell’8 settembre 2020. *Sei ancora spaventato, umano?*, si chiedeva una persona in carne e ossa facendo il titolo di questo articolo interamente scritto da un’intelligenza artificiale. La domanda va presa sul serio, e forse val la pena provare a fornire qualche riflessione per una risposta. Partendo un po’ da lontano – ovvero intorno alla metà del Quattrocento (1448-1454), con Gutenberg.

L’eccezionale rivoluzione alla quale stiamo assistendo ha infatti pochi, rarissimi antecedenti nella storia – pari soltanto, per la forza degli effetti, all’invenzione del libro. Bacone sosteneva che l’arte della stampa (insieme alla polvere da sparo e alla bussola) aveva sconvolto l’assetto del mondo, per sempre. Anche Marx sosteneva la stessa identica tesi, riprendendola fra l’altro da Hegel: “La polvere da sparo, la bussola, la stampa sono le tre grandi scoperte introdotte dalla società borghese. La polvere da sparo dissolve la cavalleria, la bussola apre il mercato mondiale e crea le colonie – mentre la stampa diventa lo strumento del

protestantesimo e in generale del risveglio della scienza: la più importante leva per costituire i presupposti di un indispensabile sviluppo spirituale” (Marx 1980, 93).

Ora, gli aspetti più tecnici, sociali, politici di quanto sta avvenendo al giorno d’oggi con la rivoluzione digitale sono evidenti. Ma sono gli effetti a esser ancora da interrogare. Può essere utile in questo senso stabilire un’analogia con qualcosa di simile accaduto nel passato, in particolare con la rivoluzione a partire da Gutenberg. Dopo la comparsa delle prime botteghe dei tipografi e la nascita del libro, la sua diffusione e l’evoluzione delle tecniche, iniziò una rivoluzione a lungo silenziosa e inavvertita che provocò nelle teste e nelle cose, negli individui e nella società, negli umili e nei potenti, cambiamenti decisivi e irreversibili. Esattamente quel che sta succedendo oggi: però faticiamo a mettere a fuoco, distintamente – come del resto allora non venne subito colto il *mondo nuovo* frutto della rivoluzione del libro, i cui risultati saranno nientemeno che Rinascimento, Riforma, trasformazione del *libro della natura* – ovvero la scienza nuova. Ma insieme anche un paio di secoli di instabilità, guerre e persecuzioni religiose.

Ma cosa sappiamo di questo, ultimo mondo nuovo che sta arrivando, oggi? Quella radicale trasformazione a cui assistiamo – che a differenza di allora sta avvenendo con una velocità mai vista né immaginata – riguarda le nostre vite, che stanno cambiando in maniera sempre più rapida: il digitale influenza la formazione dell’immaginario collettivo, modifica i valori, i simboli, le icone, i miti della contemporaneità. Ma né la classe politica né tanto meno l’opinione pubblica sembrano avere gli strumenti culturali per capire e regolare questi cambiamenti. Ecco allora che questo numero dedicato allo *Human-centric approach to Artificial Intelligence* mira a raccogliere studi e riflessioni sullo sviluppo tecnologico, in particolare sulle potenzialità e sui rischi dell’intelligenza artificiale, intesa come l’insieme dei risultati frutto dell’uso della “scienza e

ingegneria della costruzione di macchine intelligenti” – iniziata nel 1956 con l’invenzione dei robot, macchine biosintetiche usate come forza lavoro, la cui denominazione discende dalla parola in lingua ceca *robota* (McCarthy 2007; Hamet e Tremblay 2017).

L’intelligenza artificiale rappresenta un concetto molto generico e ampio che ricomprende diverse tecnologie specifiche, come l’apprendimento automatico e quello profondo, le reti neurali applicate a diversi ambiti di ricerca, sperimentazione e attività scientifiche (dalla medicina all’educazione, dall’ingegneria alla finanza e così via). Le recenti discussioni sulla possibilità di un comportamento in qualche modo inaspettato o “emergente” hanno in realtà a che vedere soprattutto con l’intelligenza artificiale generativa, con l’apprendimento profondo nelle reti neurali (ispirate alla struttura del cervello umano), poiché queste si riferiscono a strutture generiche che non incorporano regole prestabilite o alberi decisionali (o altri strumenti come gli algoritmi genetici), ma devono essere addestrate su serie di dati coerenti per produrre alla fine risultati adeguati (Higuera 2023). La probabilità che informazioni incomplete e distorte, *fake news* e disinformazione possano avere un ritorno, producendo un effetto di “ingrandimento” attraverso l’intelligenza artificiale generativa, può essere imprevedibile. C’è chi, come lo storico Yuval Noah Harari, ritiene che questa tecnologia costituisca un pericolo sociale e politico, poiché appropriandosi del linguaggio, manipolandolo e sottoponendolo ad hackeraggio, potrebbe portare alla distruzione della democrazia e della civiltà umana (Harari 2023; Beccaria 2023).

Il discorso pubblico e istituzionale su questi temi pone importanti interrogativi alla normativa nazionale e sovranazionale riaccendendo il dibattito tra esigenze di sicurezza preventiva, diritto alla tutela della *privacy*, e rischi di sorveglianza pubblica di massa.

Indipendentemente dalle previsioni più o meno allarmistiche, risulta urgente procedere alla regolamentazione dell’intelli-

genza artificiale in modo che sia strettamente legata a una corretta regolamentazione dei dati e della loro governance, così come al concetto di responsabilità per garantire i diritti umani (Peyron 2021). Nel 2019 le accademie dei paesi del G7, le cui delegazioni erano riunite a Parigi, hanno firmato all'unanimità una dichiarazione su intelligenza artificiale e società (2019 *G7 Science Academies Declaration: Artificial Intelligence and Society* – si veda anche lo studio OECD 2019). Se diverse iniziative internazionali sono in corso per un confronto sugli sviluppi per orientare l'intelligenza artificiale verso il bene comune (come ad esempio la piattaforma digitale *AI for Good* dell'Unione internazionale delle telecomunicazioni delle Nazioni Unite <https://aiforgood.itu.int/>), una regolamentazione a livello globale è ancora lontana e una prima Dichiarazione comune internazionale sulla sicurezza dell'intelligenza artificiale è stata firmata a Londra nel novembre 2023 (*AI Safety Summit* 2023). Contemporaneamente la prima Commissione dell'Assemblea generale delle Nazioni Unite ha adottato la prima risoluzione sulle “armi autonome” che chiede al Segretario Generale delle Nazioni Unite di presentare un rapporto sull'argomento, tenendo conto delle preoccupazioni sul rapido sviluppo delle armi robot grazie all'uso di intelligenza artificiale nel settore militare. Da qui la proposta di mobilitazione e di impegno della comunità scientifica e della società civile per la pace (sugli accordi internazionali per il controllo e la messa al bando di queste armi si veda Farruggia 2023). L'Unione Europea, dopo essersi già dotata di normative sui servizi e sui mercati digitali, sui dati e sulla governance dei dati, ha raggiunto un accordo nel dicembre 2023 sulla prima legge al mondo denominata *AI Act*, che ambirebbe a proporsi come modello per una futura regolamentazione globale. Tuttavia, questo regolamento dell'UE che entrerà gradualmente in vigore (European Parliament, 2023) ha già ricevuto critiche da organizzazioni non governative operanti a difesa dei diritti umani,

come Amnesty international (2023) poiché, pur ponendo limiti stringenti ai “modelli fondativi” – ossia all’intelligenza artificiale generativa per testi (come Chat Gpt) o per immagini – considerati ad alto rischio e quindi in grado di causare danni sociali – non stabilisce un divieto incondizionato dei sistemi di riconoscimento biometrico delle persone in tempo reale e a distanza negli spazi pubblici. Ciò significa che viene consentito, per motivi di necessità e sicurezza nazionale, il riconoscimento facciale in caso di prevenzione di un attacco terroristico, di localizzazione di vittime o sospetti e di alcuni crimini gravi predefiniti (European Parliament 2023, Bertuzzi 2023). Il Parlamento Europeo si è impegnato a monitorare l’attuazione del programma Horizon dell’UE, che prevede anche agevolazioni e finanziamenti per la ricerca sull’intelligenza artificiale. Il programma politico *Percorso per il decennio digitale*, che definisce gli obiettivi per il 2030, auspica che il 75% delle aziende dell’UE utilizzi il cloud, l’intelligenza artificiale e i Big Data entro quell’anno, e di definire le politiche in settori come l’economia creativa, l’istruzione, la salute e molti altri ambiti industriali, sociali e culturali. Ma nonostante la sua impostazione fosse rigorosa, l’impatto è stato attenuato a causa dell’influenza di importanti industrie del settore (come la francese Mistral AI o la tedesca Aleph Alpha) esercitata sui governi rappresentati nel Consiglio dell’UE. L’esito di questo confronto pone importanti interrogativi alla normativa nazionale e sovranazionale riaccendendo il dibattito tra esigenze di sicurezza preventiva e diritto alla tutela della privacy, e rischi di sorveglianza pubblica di massa.

A partire da tali premesse il volume *Human-centric approach to Artificial Intelligence* intende stimolare una riflessione su sfide e rischi degli sviluppi dell’intelligenza artificiale, partendo da una domanda centrale che guida il ragionamento su questi temi: *quali limiti per una sua sostenibilità ecologica e umana?* La comunità accademica è stata invitata a presentare contributi tesi a indagare,

in chiave inter, multi e transdisciplinare le implicazioni etiche, sociali, culturali ed economiche per la sicurezza umana, correlate agli sviluppi e alle applicazioni di questa tecnologia, nonché le trasformazioni e gli impatti sulla società, in particolare sulle istituzioni educative, in funzione della salvaguardia della libertà e del benessere delle persone e delle comunità. Con questo contributo si intende concorrere alla riflessione e all'elaborazione di percorsi e di azioni rispondenti a una prospettiva umano-centrica, volta a ridurre il divario tra i valori degli sviluppatori essenzialmente attraverso l'utilizzo di algoritmi e quelli delle persone che li utilizzeranno o che ne saranno influenzate (Christian 2020), concentrandosi sui risultati ottenuti e dando priorità ai benefici o ai danni per gli uomini e le comunità (Shneiderman, 2022). Appare urgente inoltre integrare tale prospettiva con un'appropriata consapevolezza ecologica ed ecosistemica, fondata sulle interrelazioni e le interconnessioni tra le comunità umane e gli ecosistemi che ne garantiscono la vita, adottando quindi una prospettiva sistemica e un pensiero complesso, in grado cioè di prendere in esame gli effetti di reciprocità tra tutte le variabili coevolutive.

I saggi contenuti in questo numero intendono offrire analisi e proposte che possano orientare verso processi di condivisione ed elaborazione di buone pratiche, grazie al concorso di ricerche scientifiche provenienti da diversi ambiti: dalla filosofia alla sociologia, dalla storia alla scienza politica, dal diritto all'economia, dalle scienze dell'educazione alla psicologia, dalla medicina alle scienze della vita, dall'antropologia all'etica e alle religioni.

Considerati i notevoli e sempre più rapidi avanzamenti tecnologici che producono impatti crescenti e con effetti ambivalenti sulle attività umane e sulla vita personale, in larga parte ancora non conosciuti o sufficientemente studiati, la necessità di dialogo e di confronto sugli sviluppi e sugli indirizzi e i limiti da definire all'intelligenza artificiale costituiscono una priorità evidente. Infatti,

la sua concezione e applicazione andrebbero orientate e disciplinate in maniera rigorosa affinché non si prestino allo sviluppo di strumenti capaci di aumentare violenza, discriminazioni e distruzione di beni comuni, ma possano contribuire al benessere e alla cura delle persone, delle comunità, alla promozione della giustizia e della pace, alla salvaguardia della Terra come casa comune.

Nella letteratura scientifica e a livello di istituti e centri di ricerca (per esempio lo *Stanford Institute for Human-Centered Artificial Intelligence* - <https://hai.stanford.edu/> e lo *Human-Computer Interaction Lab* dell'Università del Maryland <https://hcil.umd.edu/> e moltissimi altri oggi diffusi in varie università e centri di ricerca nel mondo) il dibattito sull'approccio umano-centrico è iniziato da diversi anni ed è alquanto vivo con spunti critici e proposte differenti su valori, potenzialità e rischi.

Particolarmente insidioso è il paradigma tecnocratico che consiste nel pensare “come se la realtà, il bene e la verità sbocciassero spontaneamente dal potere stesso della tecnologia e dell'economia” (Francesco, 2015: § 105) e che porta a perseguire una crescita illimitata, incompatibile con la limitatezza e la preservazione delle risorse naturali. Come ulteriore sviluppo negativo di questo paradigma è stato rilevato come “l'intelligenza artificiale e i recenti sviluppi tecnologici si basano sull'idea di un essere umano senza limiti, le cui capacità e possibilità si potrebbero estendere all'infinito grazie alla tecnologia” (Francesco, 2023: § 21) e un forte allarme etico è emerso per l'aumento di potere concentrato in poche mani e in quelle organizzazioni che detengono la conoscenza e il potere economico di sfruttarla. Tuttavia, “non ogni aumento di potere è un progresso per l'umanità. Basti pensare alle tecnologie ‘mirabili’ che furono utilizzate per decimare popolazioni, lanciare bombe atomiche, annientare gruppi etnici. Vi sono stati momenti della storia in cui l'ammirazione per il progresso non ci ha

permesso di vedere l'orrore dei suoi effetti" (Francesco, 2023: § 24).

Accennando a diverse prospettive filosofiche, mentre il dualismo tra oggetti e valori, dall'impostazione cartesiana in avanti, si fonda sulla pretesa autonomia e neutralità della scienza, fondata su dati empirici, unità di misura specifiche e una logica razionale, un altro approccio, olistico e sistemico di ricomposizione delle interconnessioni tra l'uomo e la natura, constatando la presenza di valori umani in qualsiasi attività, afferma che questa associazione tra oggetti e valori sia alla base della loro creazione e del loro utilizzo, nonostante il dualismo e la focalizzazione tecnologica si sforzino di nascondere il legame degli algoritmi con l'azione umana (Taylor, O'Dell, Murphy, 2024). Partendo dalla critica al dualismo, per un'intelligenza artificiale volta al benessere di tutti gli stakeholders e di tutte le comunità umane sono stati proposti approcci comunitario-centrici come quello della filosofia africana dell'Ubuntu, fondato sull'interdipendenza e la cooperazione di tutti (Mugumbate, Nyanguru, 2013; Langat, Mwakio, Ayuku, 2020; Mhlambi, 2020; Gwagwa, Hilliard, 2022) e della "massima partecipazione possibile" (*maximum feasible participation*) volte a minimizzare discriminazioni dirette o indirette e a favorire inclusività, condivisione, benessere e resilienza (Taylor, O'Dell, Murphy, 2024).

A partire dal panorama fin qui descritto, i contributi raccolti intendono sollecitare la riflessione scientifica su diverse declinazioni e applicazioni dell'intelligenza artificiale (Benanti, 2022; Caporale, Palazzani, 2023) dalle origini storiche e filosofiche fino alle possibili modalità di sviluppo, compatibili con un approccio umanocentrico condiviso nel quadro di un ampio dibattito etico (Floridi, 2021, 2022; Ammanath, 2022). Sono state anche avanzate come opzioni da valutare eventuali autolimitazioni e moratorie al fine di assumere decisioni ponderate e responsabili, anche in base al principio di precauzione – tenendo conto anche delle emissioni climalteranti del

settore dell'innovazione digitale e tecnologica stimate pari a circa il 3-4% di quelle globali e al fatto che i *data center* utilizzano grandi quantità d'acqua per il raffreddamento (UNEP, 2022). Se da un lato non va dimenticato l'uso dell'intelligenza artificiale nel campo della medicina (ad esempio con importanti sviluppi per la prevenzione di malattie oncologiche tramite mappatura con un algoritmo – Vitali, 2023), bisogna tener presente anche l'impatto nel mondo del lavoro (una recente ricerca di Goldman Sachs prevede una forte riduzione di ore lavorative e la perdita di posti di lavoro in circa 2/3 delle mansioni esistenti - Mazza, 2023), nel sistema educativo, nella sfera privata e relazionale, in riferimento al mutamento degli stili di vita, a una visione olistica di empowerment di comunità, alle sfide per un'*ecologia integrale*, tesa a rimettere la persona umana al centro degli interessi collettivi. Numerose sono le questioni rilevanti a diversi livelli che implicano anche riflessioni critiche: il pregiudizio algoritmico (Ferrer, v. Nuenen, Such, Coté, Criado, 2021); la mancanza di responsabilità per i risultati generati dagli algoritmi "black-box", simili a scatole nere i cui meccanismi di funzionamento interni non sono conoscibili o visibili (Martin, 2019; Capogna, 2023); l'approvvigionamento non etico di minerali rari utilizzati nei dispositivi alimentati dall'intelligenza artificiale (Crawford, Joler, 2018); l'impronta ecologica dei *data center* (Crawford, 2018); lo sfruttamento del lavoro digitale coinvolto nell'etichettatura dei dati per il training e nella moderazione dei contenuti (Crawford 2021); la manipolazione algoritmica delle preferenze di consumo e di voto degli utenti (Zuboff 2019).

Marco Filoni, Filippo Maria Giordano, Giorgio Grimaldi

Bibliografia

AI Safety Summit (2023), The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023, Policy Paper, <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023> [Accessed 10 Dec. 2023].

Ammanath, B. (2022) *Trustworthy AI. A business guide for navigating trust and ethics in AI*, Hoboken, New Jersey: Wiley.

Amnesty International (2023) EU: Bloc's decision to not ban public mass surveillance in AI Act sets a devastating global precedent, 9 December 2023, <https://www.amnesty.org/en/latest/news/2023/12/eu-blocs-decision-to-not-ban-public-mass-surveillance-in-ai-act-sets-a-devastating-global-precedent/>.

Beccaria, G. (2023) Yuval Noah Harari: "Se trattiamo l'intelligenza artificiale come un oracolo rischiamo la fine della storia umana", "La Stampa", 3 maggio 2023 https://www.lastampa.it/audio/audioarticoli/2023/05/03/playlist/gli_audioarticoli_del_3_maggio_2023-12784387/.

Benanti, P. (2022) *Human in the loop. Decisioni umane e intelligenze artificiali*, Mondadori Università, Milano.

Bertuzzi, L. (2023) L'Unione europea trova la quadratura del cerchio con il primo regolamento sull'IA al mondo, "Euractiv Italia", 11 dicembre, <https://euractiv.it/section/digitale/news/lunione-europea-trova-la-quadratura-del-cerchio-con-il-primo-regolamento-sullia-al-mondo/>.

Bryson J.J., Theodorou A. (2019) How society can maintain humancentric artificial intelligence. In: Toivonen M, Saari E (eds) *Human-centered digitalization and services, translational systems sciences*. Springer, Singapore, pp 305–323.

Capogna, S. (2023). Sociology between Big data and Research Frontiers, a Challenge for Educational Policies and Skills *Quality & Quantity*, 57, 193-212, <https://doi.org/10.1007/s11135-022-01351-7>.

Caporale C., Palazzani L. (a cura di) (2023), *L'Intelligenza Artificiale: distingue frequenter. Come giungere a una comunanza etica della società del pluralismo*. Prefazione di Giuliano Amato. Consulta Scientifica del Cortile dei Gentili. Roma: Cnr Edizioni. Open access: <https://www.cortiledeigentili.com/wp-content/uploads/2023/05/Consulta-scientifica-Cortile-dei-Gentili-Volume-Intelligenza-Artificiale.pdf> [Accessed 10 Dec. 2023].

Christian, B. (2020) *The Alignment Problem: Machine Learning and Human Values*. New York: Norton.

Council of the EU (2023). Artificial intelligence act: Council and Parliament strike a deal on the first rules for AI in the world, Press release 986/23, 9 December 2023 <https://www.consilium.europa.eu/it/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/> [Accessed 10 Dec. 2023].

Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, New Haven: Yale University Press <https://doi.org/10.12987/9780300252392>.

Crawford, K., Joler, V. (2018). "Anatomy of an AI System: The Amazon Echo As An Anatomical Map of Human Labor, Data and Planetary Resources," AI Now Institute and Share Lab, (September 7, 2018) <https://anatomyof.ai> [Accessed 10 Dec. 2023].

European Parliament (2023) Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI, Press Release, 9 December 2023 <https://www.europarl.europa.eu/news/en/press-room/2023/1206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai> [Accessed 10 Dec. 2023].

Farruggia, F. (a cura di) (2023). *Dai droni alle armi autonome. Lasciare l'Apocalisse alle macchine?* Prefazione di Giorgio Parisi. Milano: FrancoAngeli. Open access: <https://series.francoangeli.it/index.php/oa/catalog/book/948> [Accessed 10 Dec. 2023].

Ferrer, X., v. Nuenen, T., Such, J.M., Coté, M., Criado, N. (2021). "Bias and Discrimination in AI: A Cross-Disciplinary Perspective," in *IEEE Technology and Society Magazine*, 40(2), 72-80, doi: 10.1109/MTS.2021.3056293.

Floridi, L. (ed.) (2021). *Ethics, Governance, and Policies in Artificial Intelligence*, Cham: Springer.

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

Francesco (2015). *Laudato si. Lettera enciclica sulla cura della casa comune*, Libreria Editrice Vaticana, Roma https://www.vatican.va/content/francesco/it/encyclicals/documents/papa-francesco_20150524_enciclica-laudato-si.html

Francesco (2023). *Esortazione apostolica Laudate Deum del Santo Padre Francesco a tutte le persone di buona volontà sulla crisi climatica*, Libreria Editrice Vaticana, Roma https://www.vatican.va/content/francesco/it/apost_exhortations/documents/20231004-laudate-deum.html.

GPT-3, “A robot wrote this entire article. Are you scared yet, human?”, *The Guardian*, Sept. 8, 2020.

Gwagwa A, Kazim E, Hilliard A (2022). The role of the African value of Ubuntu in global AI inclusion discourse: a normative ethics perspective. *Patterns. Perspective*, 3(4), April <https://doi.org/10.1016/j.patter.2022.100462> [Accessed 10 Dec. 2023].

G7 France, 2019 G7 Science Academies Declaration: Artificial Intelligence and Society, Summit of the G7 Science Academies, Paris, March 25-26, 2019, <https://royalsociety.org/-/media/about-us/international/g-science-statements/2019-g7-declaration-artificial-intelligence-and-society.pdf?la=en-GB&hash=0F5AC7386F43088FF9A5E55BBB3E56BE> [Accessed 10 Dec. 2023].

Hamet, P., Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism* 69: 36-40. <https://doi.org/10.1016/j.metabol.2017.01.011>.

Harai, Y.N., Yuval Noah Harari argues that AI has hacked the operating system of human civilisation, *The Economist*, 28 April 2023, <https://www.economist.com/by-invitation/2023/04/28/yuval-noah-harari-argues-that-ai-has-hacked-the-operating-system-of-human-civilisation>.

Higuera, A.C. (2023). What if generative artificial intelligence became conscious?, EPRS (European Parliamentary Research Service), At a Glance - Scientific Foresight: What if?, PE 753.162 – October 2023 [https://www.europarl.europa.eu/RegData/etudes/ATAG/2023/753162/EPRS_ATA\(2023\)753162_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/ATAG/2023/753162/EPRS_ATA(2023)753162_EN.pdf) [Accessed 10 Dec. 2023].

Langat, S.K., Mwakio, P.M., Ayuku, D. (2020). “How Africa should engage Ubuntu ethics and artificial intelligence” edited by R. Hajare, *Journal of Public Health International* 2(4), 20–25. <https://doi.org/10.14302/issn.2641-4538.jphi-20-3427>.

Martin, K. (2019). Ethical Implications and Accountability of Algorithms. *Journal of Business Ethics* 160, 835–850 <https://doi.org/10.1007/s10551-018-3921-3> [Accessed 10 Dec. 2023].

Marx, K. (1980). *Capitale e tecnologia*, manoscritti 1861-1863, a cura di P. Bolchini, Roma: Editori Riuniti.

Mazza, M.P. (2023). Con l'intelligenza artificiale addio a 300 milioni di posti di lavoro, lo studio di Goldman Sachs: quali ruoli rischiano di più, Open, 28 marzo 2023, <https://www.open.online/2023/03/28/usa-studio-goldman-sachs-ai-lavoro/>.

McCarthy, J (2007). What is artificial intelligence? Available at: <http://www-formal.stanford.edu/jmc/whatisai.pdf> [Accessed 10 Dec. 2023].

Mhlambi, S. (2020). “From Rationality to Relationality: Ubuntu as an Ethical and Human Rights Framework for Artificial Intelligence Governance.” Carr Center Discussion Paper Series (2020–009).

Mugumbate, J., Nyanguru, A. (2013) Exploring African philosophy: the value of Ubuntu in social work. *African Journal of Social Work* 3(1):82–100.

Murphy, P. (2024). *AI in the Movies*, Edinburgh: Edinburgh University Press.

OECD (2019). *Artificial Intelligence in Society*. Paris: OECD Publishing <https://doi.org/10.1787/eedfee77-en>.

Panesar, A. (2019). *Machine Learning and AI for Healthcare: Big Data for Improved Health Outcomes*. Berkeley, CA: Apress.

Peyron, L. (2021). “I diritti umani nella condizione digitale: un’opportunità di rilancio” in E. Larghero, G. Zeppegno (a cura di), *Diritti umani e dignità nell’era della globalizzazione. Percorsi bioetici*, Cantalupa (Torino): Effatà.

Raus, R., Silletti, A.M., Zollo, S.D., Humbley, J. (eds) (2022). Multilinguisme et variétés linguistiques en Europe à l’aune de l’intelligence artificielle/Multilinguismo e variazioni linguistiche in Europa nell’era dell’intelligenza artificiale/Multilingualism and Language Varieties in Europe in the Age of Artificial Intelligence, *De Europa*, Special issue 2022, <https://www.collane.unito.it/oa/items/show/132#?c=0&m=0&s=0&cv=0> [Accessed 10 Dec. 2023].

Shneiderman, B. (2022). *Human-Centric AI*. Oxford: Oxford University Press. (Book presentation <https://hcil.umd.edu/human-centered-ai/>) [Accessed 10 Dec. 2023].

Taylor, R.R., O’Dell, B. & Murphy, J.W. (2024). Human-centric AI: philosophical and community-centric considerations. *AI & Society*

39, 2417-2424 <https://doi.org/10.1007/s00146-023-01694-1>
Available at: <https://link.springer.com/article/10.1007/s00146-023-01694-1> [Accessed 10 Dec. 2023].

UNEP - United Nations Environment Programme (2022). How artificial intelligence is helping tackle environmental challenges, 7 November 2022 <https://www.unep.org/news-and-stories/story/how-artificial-intelligence-helping-tackle-environmental-challenges>.

Vitali, C. (2023). L'algoritmo "legge il futuro dei tumori. Così si cerca di prevedere il decorso delle malattie oncologiche grazie a un sistema di intelligenza artificiale sviluppato all'Università di Milano Bicocca, "Avvenire", suppl. "è vita" n. 795, 19 ottobre 2023.

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs.

RUBRICA *EDUCATION*

1. INTELLIGENZA ARTIFICIALE GENERATIVA NELLA DIDATTICA: VERSO UN APPROCCIO UMANO-CENTRICO

di Michele Baldassarre*, Francesco Pio Sarcina**, Anna Maria Cuzzi***

Abstract: L'introduzione delle tecnologie di Intelligenza Artificiale Generativa (GAI) nel campo della formazione scolastica è una delle evoluzioni più significative e dibattute degli ultimi anni. Il contributo intende fornire un'analisi critica di benefici e sfide legati all'Intelligenza Artificiale (IA), in modo da orientare educatori, *policy makers* e *stakeholders* nella presa di decisioni informate sulla GAI nel sistema educativo. La presentazione di esempi applicativi di tecnologie didattiche di IA, da utilizzare per l'*empowerment* dei processi di insegnamento e apprendimento, apre le strade a nuove modalità di interazione e supporto

* Dipartimento di Scienze della Formazione, Psicologia, Comunicazione, Università degli Studi di Bari "Aldo Moro", michele.baldassarre@uniba.it.

** Dipartimento di Scienze della Formazione, Psicologia, Comunicazione, Università degli Studi di Bari "Aldo Moro": francesco.sarcina@uniba.it.

*** Dipartimento di Scienze della Formazione, Psicologia, Comunicazione, Università degli Studi di Bari "Aldo Moro", anna.cuzzi@uniba.it.

Sebbene gli autori abbiano condiviso l'intera conduzione della ricerca *ivi* presentata e l'impostazione dell'articolo, si attribuisce a Michele Baldassarre la scrittura dei paragrafi: Introduzione; Conclusioni e prospettive future; a Francesco Pio Sarcina i paragrafi: 1. Decisioni educative e Intelligenza Artificiale: opportunità e rischi; 2. L'impatto della GAI sulla personalizzazione dell'apprendimento; 3. Applicazioni di IA per la didattica: nuovi strumenti al servizio dei docenti; ad Anna Maria Cuzzi i paragrafi: 4. Equilibrio tra tecnologia e umanità: considerazioni etiche sull'IA nella scuola; 5. Linee guida per un uso responsabile dell'IA generativa nell'istruzione.

Accettato dicembre 2024 - Pubblicato aprile 2025.

educativo, in un'ottica di inclusività e innovazione pedagogica che promuova sempre un approccio di carattere umano-centrico.

Parole chiave: Intelligenza Artificiale Generativa, apprendimento, tecnologie didattiche, formazione scolastica, insegnanti.

Abstract: The introduction of Generative Artificial Intelligence (GAI) technologies into the field of education is one of the most significant and debated developments in recent years. The contribution intends to provide a critical analysis of benefits and challenges related to Artificial Intelligence (AI) to guide educators, policymakers, and stakeholders in making informed decisions about JHA in the educational system. The presentation of application examples of AI educational technologies, to be used to empower teaching and learning processes, paves the way for new ways of interaction and educational support, from a perspective of inclusiveness and pedagogical innovation that always promotes a human-centred approach.

Keywords: Generative Artificial Intelligence, learning, didactic technologies, school education, teachers.

Introduzione

L'avvento dell'Intelligenza Artificiale (IA) ha radicalmente trasformato gli scenari di molti settori e ambiti, non ultimo quello educativo e della didattica. Volendo creare un parallelismo tra i due mondi – quello della pedagogia e quello dell'automazione – apparentemente lontani, si potrebbe paragonare la definizione di Intelligenza Artificiale a uno studente diligente: l'IA è un sistema complesso e dinamico capace di apprendere sia in modo autonomo che attraverso attività di *training* e di espandere continuamente le proprie conoscenze, grazie alla disponibilità di grandi quantità di dati che vengono sottoposti a processi di analisi e rielaborazione;

questa capacità di apprendimento consente all'IA di svolgere compiti complessi e sviluppare il ragionamento logico, la pianificazione strategica e la creatività, abilità fondamentali per risolvere svariate problematiche e utili per adattarsi a contesti nuovi e mutevoli (Russell & Norvig, 2021). Questa articolata descrizione, che sottende processi di apprendimento adattivo e di acquisizione di competenze, rispecchia gli obiettivi ideali per uno studente impegnato nel percorso educativo: partendo da un vasto sistema di conoscenze, l'allievo deve imparare autonomamente (o col supporto di strategie fornite dal docente) a sviluppare abilità di *problem solving*, *decision making*, creatività e pensiero critico, utili al raggiungimento del successo formativo e ad affrontare le sfide della società odierna. Proprio da questa affinità con il processo di apprendimento umano, l'IA trova applicazioni particolarmente efficaci nei contesti di istruzione e formazione.

L'Intelligenza Artificiale, specialmente quella generativa, offre nuove opportunità per la progettazione e la valutazione didattica: attraverso strumenti basati su sistemi di *Generative Artificial Intelligence* (GAI), i docenti possono generare contenuti educativi personalizzati, facilitare la creazione di quiz valutativi e ottenere feedback su come migliorare le lezioni (Lorenzoni, 2024). Tuttavia, è essenziale comprendere ed essere consapevoli che queste tecnologie devono essere utilizzate per potenziare – e non sostituire – l'interazione umana. L'orientamento e la prospettiva futura che la ricerca pedagogica cerca di promuovere è quella di un uso degli strumenti di IA come mezzi per arricchire l'esperienza educativa, che permettano agli insegnanti di dedicare più tempo all'interazione diretta e alla *mentorship*, due fattori fondamentali per lo sviluppo olistico e il successo scolastico degli studenti. Sostenere e valorizzare un approccio umano-centrico all'IA significa mettere al centro dell'innovazione tecnologica i bisogni, i valori e il benessere delle persone, assicurando che le tecnologie vengano

sviluppate e applicate in modo etico e responsabile. In questo modo, è possibile ottimizzare i processi di apprendimento e, allo stesso tempo, garantire che la tecnologia rimanga uno strumento al servizio dello sviluppo umano, rafforzando ulteriormente il ruolo insostituibile del docente (Corrado & Pasta, 2024).

1. Decisioni educative e Intelligenza Artificiale: quali rischi per la didattica

Quando si parla di IA si utilizzano termini che normalmente vengono attribuiti all'uomo ("intelligenza", "reti neurali", "apprendimento"), ma che in realtà sono qualificati come artificiali, alludendo al fatto che non vengono associati ad una forma vivente. Ciò deriva dal sogno utopistico di far sì che un computer potesse replicare l'intelligenza umana, potenziandola da un punto di vista delle prestazioni; tuttavia, come sostenuto da Floridi e Cabitza (2021), queste macchine sono molto abili nell'eseguire compiti, ma non sono per nulla intelligenti, al punto che bisognerebbe parlare di "agenti artificiali". La presenza e la costante evoluzione di questi strumenti, che migliorano e aumentano rapidamente la loro produttività in modo esponenziale, porta ad integrarli nella quotidianità umana in maniera quasi del tutto automatica: infatti, si potrebbe dire che «la vita, oggi, non è tanto catturata dal digitale, quanto piuttosto costruita per mezzo di essa» (Elliott, 2019:138). Questa rivoluzione ha abbracciato anche il mondo della scuola, in cui le tecnologie di Intelligenza Artificiale si pongono come strumenti di potenziamento dei processi di insegnamento-apprendimento e di supporto adeguato per docenti e studenti, purché venga sempre preservato l'approccio umano-centrico. I docenti rimangono figure centrali ed essenziali per interpretare e adattare i contenuti didattici alle necessità specifiche

di ciascun studente, un processo che richiede sensibilità, empatia e comprensione delle dinamiche emotive e sociali, aspetti che i sistemi di IA, per quanto avanzati, non possono ancora appieno replicare (Panciroli & Rivoltella, 2023). A sostegno di questa tesi, il documento di indirizzo del Ministero dell'Istruzione *Piano Nazionale Scuola Digitale* (PNSD) afferma che nessun passaggio educativo può prescindere da una reciprocità docente-discente e che la tecnologia, a sua volta, non può prescindere da questo fondamentale rapporto umano (Cuomo *et al.*, 2022). Se pensata per arricchire l'interazione umana, l'IA a scuola diventa uno strumento di *empowerment* e facilitazione per i docenti che – attraverso un uso consapevole e critico – possono essere “sollevati” da pratiche burocratiche e ripetitive e guadagnare tempo prezioso da dedicare agli aspetti educativi più profondi (come il supporto motivazionale e il confronto diretto con gli studenti). Il rapporto docente-studente, infatti, costituisce il cuore pulsante del processo educativo, che l'IA difficilmente può emulare (nonostante oggi si stia assistendo alla creazione di primi sistemi di IA in grado di sviluppare abilità sociali ed emotive). Alla luce di questo processo di integrazione dell'IA nei contesti di istruzione, appare chiaro che il tema dell'automazione sulle decisioni educative susciti molto interesse e altrettante preoccupazioni. Benché *chatbot* e applicazioni di IA siano in grado di automatizzare alcuni processi di personalizzazione dei percorsi di apprendimento e di valutazione di conoscenze degli studenti, ci sono decisioni critiche che richiedono la competenza e la sensibilità di un docente umano, specialmente quando si tratta di interpretare risultati in un contesto complesso e multifattoriale e di considerare aspetti emotivi e sociali profondi (Ranieri, 2022). La valutazione delle competenze socio-emotive e la gestione delle dinamiche relazionali in classe sono ambiti in cui il giudizio umano è fondamentale per cogliere sfumature motivazionali, resilienza e autostima, elementi

che sfuggono dal controllo dei modelli algoritmici; pertanto, l'eccessivo affidamento all'IA, per analizzare e dare feedback su tali aspetti, potrebbe condurre a risultati incompleti o distorti. Ulteriori rischi in cui si potrebbe incorrere delegando totalmente i processi decisionali educativi alle macchine riguardano i pregiudizi algoritmici e l'eccessiva interpretazione meccanica dei dati (Russell & Norvig, 2021): l'IA interpreta le informazioni in modo rigorosamente logico e quantitativo, spesso senza considerare che i *dataset* potrebbero contenere *bias* discriminativi. Ad esempio, uno studente che presenta difficoltà temporanee potrebbe essere valutato dall'IA come “non performante”, tralasciando possibili ragioni legate a difficoltà personali o situazioni contingenti; questa prospettiva strettamente algoritmica rischia di etichettare e limitare gli studenti, piuttosto che offrire un supporto personalizzato che tenga conto della loro unicità e delle loro necessità specifiche. Al fine di mitigare le conseguenze negative che ne derivano, si evidenzia, quindi, la necessità di un bilanciamento adeguato tra automazione e controllo umano (Floridi, 2022): tale approccio misto consente di sfruttare la precisione e la velocità dell'IA, mantenendo però un “filtro umano” che garantisce decisioni pedagogiche sensibili e orientate al benessere degli studenti.

2. L'impatto della General Artificial Intelligence sulla personalizzazione dell'apprendimento

L'Intelligenza Artificiale (IA) sta modificando radicalmente il panorama educativo, introducendo nuovi modelli di apprendimento che pongono al centro il singolo studente e le sue esigenze specifiche (Badino *et al.*, 2023). La personalizzazione si concretizza attraverso piattaforme di apprendimento adattivo e sistemi di

tutoraggio virtuale che adeguano in tempo reale i contenuti e le modalità di insegnamento in base ai bisogni e ai ritmi individuali di ogni studente, contribuendo a una partecipazione più attiva e coinvolgente (Panciroli *et al.*, 2020). L'IA, infatti, è in grado sia di identificare le aree di forza e di miglioramento, che di monitorare l'andamento e le performance degli studenti, suggerendo attività o risorse per colmare lacune o consolidare competenze. Ad esempio, i tutor virtuali consentono agli studenti di avanzare nel loro percorso formativo con tempi personalizzati e di affrontare concetti complessi con il supporto di feedback immediati che incoraggiano l'autocorrezione, il miglioramento continuo e la motivazione degli studenti (Ciasullo, 2023). Nonostante i numerosi vantaggi derivanti da questo approccio è tuttavia essenziale che questa personalizzazione non avvenga in modo meccanico: il ruolo del docente rimane fondamentale per interpretare i risultati e orientare le scelte educative affinché siano coerenti con il contesto e le caratteristiche personali degli studenti (Badino *et al.*, 2023). Integrando l'IA con la *mentorship* educativa degli insegnanti, infatti, l'educazione diventa non solo più avanzata, ma anche più umana, mantenendo l'alunno al centro del processo formativo. L'Intelligenza Artificiale apre a nuove prospettive nella creazione di ambienti di apprendimento dinamici e collaborativi, in cui gli studenti possono essere attivamente coinvolti in discussioni e progetti di gruppo, grazie a strumenti che facilitano la condivisione e la collaborazione (Ferrarese, 2024). Gli algoritmi di cui si serve l'IA monitorano le interazioni tra studenti, suggerendo sinergie e adattando i gruppi in base a competenze complementari; inoltre, questi ambienti virtuali consentono agli studenti di interagire e vivere esperienze in contesti reali ma simulati, favorendo l'applicazione teorica in contesti pratici (come in un laboratorio virtuale di scienze o un viaggio nel passato per studiare la storia). Queste simulazioni non solo coinvolgono maggiormente gli studenti, ma contribuiscono a una comprensione

più profonda e consolidata delle materie studiate. L'apprendimento attraverso ambienti di apprendimento virtuali migliora quindi le capacità di *problem solving* e l'adattamento a situazioni nuove, allineandosi con i principi del Costruttivismo, secondo cui gli studenti sono più motivati e tendono a ricordare meglio ciò che imparano quando possono applicare immediatamente le conoscenze in contesti simili alla realtà (Jonassen *et al.*, 1998). L'uso dell'IA nella creazione di scenari personalizzati permette di simulare situazioni che sarebbero difficili da ricreare in un'aula tradizionale, realizzando esperienze didattiche uniche. Questa metodologia, integrata nelle tecniche di *smart learning design* (Sancassani & Mariconda, 2023) permette di trasformare l'insegnamento in un processo dinamico e applicabile alle esigenze specifiche di ogni studente, facendo dell'IA un facilitatore in grado di offrire risposte e adattamenti continui alle interazioni degli utenti. Un ulteriore beneficio degli scenari creati dall'IA è il potenziamento dell'autonomia: lavorando in un contesto applicativo, gli studenti possono esplorare diversi approcci, prendere decisioni e ricevere feedback immediati dalle loro azioni (Di Bello, 2023). Questa tipologia di apprendimento, basata sull'esperienza diretta e sull'esplorazione, incoraggia lo sviluppo di competenze di auto-regolazione e auto-valutazione, in cui lo studente diventa soggetto attivo del proprio percorso di crescita (Pastorelli, 2024).

La combinazione sinergica tra tutoraggio virtuale, ambienti collaborativi e supporto adattivo trasforma radicalmente l'esperienza educativa, rendendola più inclusiva e rispondente alle caratteristiche di ogni discente (Badino *et al.*, 2023). Tuttavia, promuovere questa inclusività comporta maggiore attenzione affinché si eviti la creazione di disuguaglianze e disparità, in quanto la disponibilità e l'accesso agli strumenti di IA variano notevolmente tra istituzioni educative. È quindi fondamentale che le politiche educative promuovano l'equità nell'accesso alle tecnologie, affinché

il divario digitale non si traduca anche in un divario formativo (Pastorelli, 2024).

3. Applicazioni di IA per la didattica: nuovi strumenti al servizio dei docenti

Le applicazioni basate sull'intelligenza artificiale (IA) rappresentano strumenti innovativi che non solo supportano, ma arricchiscono e personalizzano l'esperienza di apprendimento, stimolando la creatività e l'espressione individuale degli studenti (Di Bello, 2023). Il modello *Technological Pedagogical Content Knowledge* (TPACK), elaborato da Mishra e Koehler (2006) della Michigan University, rappresenta un quadro concettuale fondamentale per integrare efficacemente la tecnologia nell'insegnamento. Il modello suggerisce che un'efficace integrazione dell'IA in ambito educativo richiede un equilibrio tra conoscenza del contenuto (*Content Knowledge*), conoscenza pedagogica (*Pedagogical Knowledge*) e competenze tecnologiche (*Technological Knowledge*). È evidente come, grazie alla tecnologia, sia possibile affrontare una sfida comune a tutti i docenti, ovvero quella di mantenere alta l'attenzione degli studenti e di potenziare il loro coinvolgimento. Questo concetto è ben espresso nelle parole di Recalcati (2014: 99), nelle quali molti insegnanti spesso si riconoscono: «È ancora oggi quello a cui personalmente punto ogniqualvolta mi trovo in un'aula: tenere sveglio chi mi ascolta, impedire che la sua testa cada comatosa sul banco, forzare la tendenza al sonno, provocare risvegli, far sentire la forza della parola».

Oggi, i docenti dispongono di nuovi strumenti per affrontare questa sfida: tecnologie come la Realtà Aumentata e Virtuale (RA e RV), le piattaforme interattive e l'Intelligenza Artificiale permettono di trasformare le lezioni in esperienze

dinamiche, immersive e personalizzate (Biaggio, 2022). Tuttavia, è opportuno ricordare che la tecnologia è uno *strumento* e non un *fine*. La sua efficacia dipende dalla capacità del docente di utilizzarla in modo consapevole, per facilitare una partecipazione attiva senza compromettere la qualità del contenuto. In questo senso, la *forza della parola* di cui parla Recalcati resta fondamentale: la tecnologia può amplificare questa forza, ma non può sostituirla. Il docente ha il compito di mantenere viva quella scintilla di dialogo e relazione autentica che stimola un apprendimento vero e profondo, trasformando l'aula in un luogo di risvegli continui. La Tabella 1 intende suggerire alcuni siti e applicazioni di IA generativa, utili per la preparazione, lo svolgimento e la valutazione delle attività didattiche.

Tabella 1: Siti e applicazioni di IA generativa per la didattica

Logo	Nome e sito App	Descrizione
	Gamma App (https://gamma.app/)	Piattaforma di presentazione visiva che crea presentazioni e slide in modo automatico. L'utente fornisce il testo e l'IA organizza le diapositive, ottimizzando il <i>design</i> e la struttura dei contenuti. Nella didattica, facilita la creazione di presentazioni per lezioni interattive e visive, supportando studenti con necessità di supporti visivi.
	Leonardo AI (https://leonardo.ai/)	Piattaforma per la grafica generativa, che crea contenuti visivi e grafici da descrizioni testuali. L'utente fornisce <i>prompt</i> testuali e l'IA genera immagini artistiche dettagliate. Nella didattica, è utile per illustrare concetti complessi e creare materiali visivi che stimolano la creatività degli studenti.

	Genially (https://genial.ly/)	Piattaforma utile a creare presentazioni, infografiche e contenuti interattivi, come <i>escape room</i> digitali. Utilizza modelli personalizzabili e suggerimenti di <i>design</i> per creare contenuti coinvolgenti. È uno strumento versatile per realizzare lezioni dinamiche e interattive, incoraggiando la partecipazione degli studenti.
	Canvas AI (https://canvas.co/canvas-ai)	Editor grafico basato sull'IA che facilita la creazione di <i>layout</i> grafici e contenuti visivi. Grazie agli strumenti guidati dall'IA, l'utente può creare risorse esteticamente coerenti creando risorse visive accattivanti per supportare lezioni e presentazioni.
	Crayon (https://www.crayon.com/)	<i>Tool</i> di disegno digitale che genera immagini a partire da descrizioni testuali inserite dall'utente. L'IA converte il testo in disegni, visualizzando concetti complessi.
	Algor (https://www.algor-education.com/)	Strumento educativo per la creazione di mappe concettuali interattive. Organizza automaticamente le informazioni in mappe, facilitando la comprensione e l'organizzazione concettuale, supportando gli studenti nel comprendere e visualizzare relazioni tra concetti in modo strutturato.
	D-ID (https://www.d-id.com/)	Piattaforma per creare video animati e interattivi con avatar realistici. Gli utenti possono inserire anche un testo per animare gli avatar, che comunicano in modo naturale grazie all'IA. Nella

		didattica, è utile per spiegare concetti attraverso avatar interattivi, rendendo i contenuti più dinamici.
	Diffit (https://diffit.me)	Strumento utile per generare attività didattiche e materiali di studio a partire da un argomento. L'IA automatizza la creazione di risorse per l'apprendimento preparando rapidamente materiali per lezioni o compiti, supportando la personalizzazione dei contenuti.
	Hey Gen (https://www.heygen.com/)	Generatore di video animati che crea contenuti video a partire da testo. L'utente inserisce descrizioni e l'IA genera video favorendo un apprendimento attivo.
	Magic School (https://magicschool.ai/)	Web app progettata per supportare insegnanti e educatori nella creazione di materiali didattici. Offre strumenti e contenuti per l'insegnamento con risorse automatizzate e materiali interattivi per varie materie.
	Magic Slides (https://magicslides.app/)	Generatore di slide automatico che crea presentazioni su un dato argomento. Inserendo un argomento, l'IA genera le slide. Nella didattica, facilita la preparazione di lezioni, risparmiando tempo nella creazione di materiali didattici visivi.
	ChatGPT (https://chat.openai.com/)	Modello di linguaggio di IA che risponde a domande e fornisce assistenza testuale su vari argomenti. Nella didattica, supporta studenti e insegnanti nella ricerca, tutoraggio e sviluppo delle idee, migliorando la comprensione.

 DALL-E 3	DALL-E 3 (https://openai.com/dall-e-3/)	Generatore di immagini che crea contenuti visivi originali da descrizioni testuali, supportando la visualizzazione di concetti astratti o difficili da rappresentare e migliorando l'interazione visiva con gli studenti.

Fonte: Ricerca a cura degli autori

Volgendo uno sguardo al futuro, appare evidente come l'integrazione dell'Intelligenza Artificiale nell'educazione possa aprire nuove frontiere di insegnamento, rendendo i processi di apprendimento più adattivi e centrati sugli studenti. È fondamentale, però, promuovere una sinergia tra innovazione tecnologica e pedagogia, investendo in ricerca e sviluppo per realizzare strumenti che non perdano di vista gli aspetti etici e la prospettiva di un approccio umano-centrico.

4. Equilibrio tra tecnologia e umanità: considerazioni etiche sull'IA nella scuola

L'integrazione dell'Intelligenza Artificiale nel contesto scolastico comporta una riflessione approfondita sulle implicazioni etiche e sulle percezioni sociali, che risentono sia della natura propria di questi sistemi che del loro uso in ambito educativo. Un primo aspetto è quello della poca trasparenza delle reti neurali che, soprattutto nei modelli di *Deep Learning*, possiedono un numero elevato di parametri e connessioni: la complessità strutturale e di funzionamento – si parla di *black box* (Panciroli & Rivoltella, 2023) poiché non si conosce precisamente il processo che la macchina segue per generare *output* – pone un problema considerevole nel momento in cui l'IA viene usata per prendere decisioni che hanno

un impatto sulle vite umane. L'utilizzo dell'IA generativa nella scuola risente dell'assenza di questa trasparenza, in quanto modelli conversazionali come *ChatGPT*, *Gemini*, *Copilot* – che trovano grande applicazione tra studenti e docenti – non possiedono né una coscienza etica né una bussola morale: le espressioni linguistiche sono percepite come portatrici di valori e presupposti relativi alla struttura sociale (Galletti & Zipoli Caiani, 2024). Pertanto, l'aspetto sfidante consiste, da un lato, nel saper individuare allucinazioni e *bias*, dall'altro, nel saper utilizzare questi sistemi in modo critico e responsabile, tenendo conto del dualismo tra mancanza di consapevolezza e capacità di agire in modo eticamente influente che li caratterizza. La seconda questione etica sollevata dall'uso dell'IA riguarda la *privacy* e la protezione dei dati. I sistemi di IA sono in grado di raccogliere e analizzare un'enorme quantità di dati personali degli studenti, tra cui prestazioni accademiche, interazioni, abitudini di apprendimento e persino tratti comportamentali (Miceli, 2023). Questi dati, spesso sensibili, richiedono misure di protezione rigorose, poiché una gestione impropria o una fuga di dati può esporre gli studenti a rischi di violazione della *privacy* e discriminazioni sociali. Il recente *AI Act* (Regolamento europeo sull'Intelligenza Artificiale, 2024) è stato concepito anche per garantire che questi sistemi, applicati in settori delicati come quello educativo, rispettino principi di trasparenza, sicurezza e controllo, imponendo agli sviluppatori l'obbligo di documentare e spiegare chiaramente come avviene il processo di trattamento dei dati. Infine, una terza questione interessa i temi dell'equità e dell'inclusione, promossi come principi cardine all'interno della *Raccomandazione sull'Intelligenza Artificiale* dell'OCSE (2024). Infatti, l'accesso ai dispositivi e alle risorse tecnologiche necessarie per beneficiare dell'IA non è sempre garantito a tutti in modo equo soprattutto in quei contesti svantaggiati dove il divario digitale (Badino *et al.*, 2023) può portare a un diverso grado di opportunità

nell'istruzione e a una rigida stratificazione sociale. Ciò richiede che scuole, governi e sviluppatori collaborino affinché gli strumenti di IA non diventino strumenti di esclusione. Se da un lato l'IA applicata ai processi di insegnamento e apprendimento offre opportunità senza precedenti per la personalizzazione e il miglioramento delle pratiche educative, dall'altro, apre a un'attenta riflessione sulle modalità per affrontare queste sfide, che la denoterebbero come uno strumento pericoloso e dannoso per la scuola. L'equilibrio tra tecnologia e fattore umano può essere garantito solo attraverso la promozione di comportamenti e indicazioni etiche e responsabili, che incentivino un approccio integrato attento sia alle potenzialità dell'IA, sia alla componente empatica, relazionale e interpretativa (Moriggi & Pireddu, 2024).

5. Linee guida per un uso responsabile dell'IA generativa nell'istruzione

In questo paragrafo sono sintetizzate alcune strategie e misure di base (Tabella 2) che docenti, formatori e dirigenti possono adottare per comprendere in che misura l'IA e i dati prodotti dai modelli generativi possono essere usati nelle attività scolastiche, con lo scopo di migliorare i risultati degli studenti, non perdendo di vista le questioni etiche precedentemente affrontate.

Tabella 2: Strategie per un uso responsabile dell'IA a scuola

Utilizzare domande orientative	Le domande orientative sono utili per prendere in analisi un applicativo di IA prima di introdurlo in una scuola o per analizzare il suo impatto durante il suo impiego. Questi interrogativi portano i principali stakeholders a riflettere sui seguenti temi: • trasparenza (<i>Il sistema di IA è allineato agli obiettivi di apprendimento degli studenti? Quanto sono affidabili le sue previsioni, valutazioni e classificazioni nell'interpretare e giustificare l'appropriatezza del suo utilizzo in ambito educativo?</i>); • diversità, non discriminazione e equità (<i>Il sistema di IA fornisce modalità di interazione adeguate per studenti con disabilità o bisogni educativi speciali? È progettato per rispettare i discenti e adattarsi in modo efficace alle loro esigenze individuali?</i>); • sorveglianza intervento umano (<i>Il ruolo dell'insegnante è chiaramente definito per assicurare il suo coinvolgimento attivo durante l'uso del sistema di IA? In che modo l'adozione del sistema di IA influenza le funzioni didattiche e pedagogiche dell'insegnante?</i>); • benessere sociale e ambientale (<i>Il sistema di IA comunica in modo chiaro che la sua interazione sociale è una simulazione e che non è in grado di provare sentimenti o empatia?</i>); • riservatezza e governance dei dati (<i>Sono presenti meccanismi per garantire l'anonimato dei dati sensibili? Esistono procedure che limitano l'accesso ai dati esclusivamente alle persone autorizzate e che ne hanno effettiva necessità?</i>); • accountability (<i>Come vengono valutati l'efficacia e l'impatto del sistema di IA e in che modo tale valutazione tiene conto dei valori fondamentali dell'istruzione?</i>); • robustezza tecnica e sicurezza (<i>Vengono fornite informazioni adeguate per assicurare studenti e genitori sulla robustezza tecnica e sulla sicurezza del sistema?</i>).

Programmare azioni per un uso efficace dell'IA e dei dati	Gli istituti che intendono integrare nei propri processi educativi l'IA devono incoraggiare azioni collaborative e di riesame delle attività scolastiche: prevedere l'impatto che l'IA e l'uso dei dati può avere è un compito complesso, ma adottando un approccio incrementale nello sviluppo, nell'implementazione e nella valutazione di queste tecnologie è possibile tenere sotto controllo gli effetti sociali ed intervenire tempestivamente in caso si manifestino conseguenze indesiderate. Un'ulteriore azione è quella di introdurre politiche e procedure a livello scolastico per definire le aspettative e fornire orientamenti sul modo in cui affrontare in maniera coerente eventuali problemi, identificando tendenze in merito ai comportamenti da considerare opportuni, inopportuni o inaccettabili. Sperimentare il sistema con una specifica coorte di discenti prima ancora di estenderlo a tutta la scuola attraverso un progetto pilota rappresenta un'altra strategia per poter far emergere un primo giudizio informato sull'efficacia del sistema di IA, in termini di miglioramento dei risultati di apprendimento, convenienza economica e uso etico.
Sensibilizzare e coinvolgere la comunità	La collaborazione tra educatori è essenziale per migliorare la scuola e favorire il successo degli studenti, poiché permette il sostegno reciproco e una gestione condivisa delle responsabilità. Questa collaborazione promuove scelte più consapevoli e un approccio coerente nell'uso dell'IA e dei dati a livello scolastico. Inoltre, la condivisione di esperienze tra scuole facilita l'apprendimento dalle pratiche di implementazione dell'IA adottate altrove. Anche partecipare a progetti regionali, nazionali o europei offre agli educatori l'opportunità di collaborare in ricerche applicate, contribuendo a definire le basi per un futuro uso etico e sicuro dell'IA nelle scuole. È fondamentale coinvolgere genitori e studenti, spiegando in modo chiaro quali dati vengono raccolti, come vengono utilizzati e protetti, e permettendo loro di esprimere feedback e preoccupazioni. Infine, vista la rapida evoluzione dell'IA, insegnanti, educatori e

	dirigenti scolastici devono mantenersi aggiornati attraverso una formazione professionale continua, garantendo così un utilizzo responsabile e informato dell'IA in ambito educativo.
Sviluppare nuove competenze in materia di IA	Considerando la rapida trasformazione del panorama educativo, l'acquisizione e lo sviluppo di competenze in materia di IA sono diventati aspetti fondamentali per docenti e formatori. Le principali competenze digitali individuate, strettamente connesse all'IA, sono: • Comprendere i concetti fondamentali dell'IA (algoritmi, <i>Machine Learning</i>, reti neurali e modelli di <i>Deep Learning</i>). • Selezionare applicazioni di IA per l'educazione (assistenti virtuali, strumenti di valutazione automatizzata, <i>chatbot</i> per l'apprendimento). • Personalizzare l'insegnamento adattando i contenuti e le metodologie in base ai dati e alle analisi dell'IA, offrendo un apprendimento su misura per ogni studente. • Conoscere i principi etici dell'IA, le implicazioni, i rischi di <i>bias</i> algoritmici e sviluppare la consapevolezza di un rapporto responsabile con le tecnologie. • Analizzare criticamente e identificare i pregiudizi potenzialmente presenti negli <i>output</i> ottenuti dall'IA generativa, al fine ridurre i rischi di discriminazione. • Promuovere un'alfabetizzazione digitale critica degli studenti, insegnando loro i fondamenti dell'IA e promuovendo una consapevolezza circa l'impatto sociale di questa tecnologia. • Partecipare alla formazione continua, investendo il proprio tempo in percorsi di aggiornamento professionale sulle innovazioni nel campo dell'IA, al fine di rimanere sempre al passo con i rapidi cambiamenti che connotano questi sistemi.

Fonte: *Orientamenti etici per gli educatori sull'uso dell'Intelligenza Artificiale (IA) e dei dati nell'insegnamento e nell'apprendimento, 2024*

Conclusioni e prospettive future

Nel contesto educativo attuale, in cui l'avvento delle tecnologie di IA è diventato centrale nel dibattito pubblico, il ruolo del docente come guida e facilitatore si dimostra più che mai indispensabile. Pur offrendo strumenti avanzati per personalizzare e ottimizzare i percorsi di apprendimento, l'IA non può sostituirsi totalmente al contributo umano nelle fasi di progettazione, valutazione e gestione delle dinamiche di classe (Tammaro, 2024) poiché il docente rimane il vero e unico progettista di esperienze significative di insegnamento e apprendimento. Il docente assume il ruolo di mediatore tra tecnologia e alunni in quanto viene chiamato a interpretare e contestualizzare i dati ottenuti: se l'IA rileva che uno studente ha difficoltà in una determinata area di apprendimento, il docente può analizzare questi risultati alla luce di altri fattori, quali l'umore dello studente, le sue relazioni con i compagni ed eventuali problemi personali o familiari, che – in quel momento – potrebbero influenzarne il rendimento scolastico. Tale sensibilità permette di adattare gli interventi educativi alle specifiche esigenze di ciascun alunno, creando così un contesto di apprendimento sicuro e supportivo (Jaimez, 2023). La componente relazionale e affettiva che il docente instaura con gli studenti rappresenta un'altra area in cui l'IA difficilmente può intervenire e porsi come alternativa alla figura umana. Le emozioni giocano un ruolo cruciale nell'apprendimento, in quanto studenti che si sentono compresi, supportati e incoraggiati tendono a mostrare maggiore motivazione e impegno nello studio. Un altro aspetto che i sistemi di IA non sono ancora in grado di valutare pienamente è quello di saper identificare le differenze tra i vari stili di apprendimento e le diverse intelligenze degli studenti. Howard Gardner (2007), nella teoria delle intelligenze multiple, ha dimostrato come gli studenti possano apprendere attraverso

modalità molto diverse tra loro: alcuni preferiscono approcci visivi, altri linguistici, altri ancora apprendono meglio attraverso attività pratiche o interazioni sociali. Gli strumenti di IA possono adattare i contenuti, ma spetta al docente l'abilità di riconoscere quale tipo di intelligenza stimolare e come personalizzare l'approccio educativo.

Per concludere, la sfida del futuro risiede nel saper coniugare la precisione analitica dell'IA con l'empatia e la capacità intuitiva dei docenti, dando vita a una pedagogia che non perde mai di vista l'individuo. Si tratta di un approccio che vede nell'IA non un sostituto, ma uno strumento potenziato al servizio di un insegnamento che valorizzi la dimensione emotiva e relazionale, coltivando ambienti di apprendimento in cui ogni studente possa crescere non solo intellettualmente, ma anche sotto il profilo umano. Raggiungere questo equilibrio significherà dare vita a un'educazione che prepara gli studenti a vivere consapevolmente nell'odierna società tecnologica, senza rinunciare al cuore dell'esperienza educativa: la connessione autentica e l'ispirazione reciproca tra docenti e discenti.

Bibliografia

Badino, M., D'Asaro, F.A. & Pedrazzoli, F. (2023), *Educare all'IA. La sfida didattica dell'Intelligenza Artificiale: ChatGPT e Gemini*. Sanoma.

Baggio, T. (2022), Tecnologia educativa nelle lezioni online e in presenza: suggerimenti per una didattica multimediale e inclusiva. *Italiano LinguaDue*, 14(2), 274-296. <https://doi.org/10.54103/2037-3597/19662>.

Celik, I. (2023), Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior*, 138. <https://doi.org/10.1016/j.chb.2022.107468>.

Ciasullo, A. (2023), ChatGPT: Le implicazioni pedagogiche e le possibilità dell'Intelligenza Artificiale. *Research Trends in Humanities*, 10, 68-76. <https://doi.org/10.6093/2284-0184/9762>.

Commissione europea, Direzione generale dell'Istruzione, della gioventù, dello sport e della cultura (2022), *Orientamenti etici per gli educatori sull'uso dell'intelligenza artificiale (IA) e dei dati nell'insegnamento e nell'apprendimento*. Ufficio delle pubblicazioni dell'Unione europea.

Corrado, V. & Pasta, S. (Eds.). (2024). *Intelligenza artificiale e sapienza del cuore*. Brescia: Morcelliana - Scholé.

Cuomo, S., Biagini, G. & Ranieri, M. (2022), Artificial Intelligence Literacy, che cos'è e come promuoverla. Dall'analisi della letteratura ad una proposta di Framework. *Media Education. Studi, ricerche e buone pratiche*, 12 (2). <https://doi.org/10.36253/me-13374>.

Cuomo, S., Ranieri, M. & Biagini, G. (2024), *Scuola e intelligenza artificiale. Percorsi di riflessione pedagogica e didattica*. Roma: Carocci.

Di Bello, B. (2023), *Intelligenza artificiale per la scuola. Un approccio umanistico all'uso didattico dell'IA generativa*. Milano: Hoepli.

Dorsey, D.W. & Michaels, H.R. (2022), Validity arguments meet artificial intelligence in innovative educational assessment. *Journal of Educational Measurement*, 59(3), 267–271. <https://doi.org/10.1111/jedm.12330>.

Elliott, A. (2019), *La cultura dell'Intelligenza Artificiale. Vita quotidiana e rivoluzione digitale*. Torino: Codice.

Ferrarese, G. (2024), *101 idee per usare l'Intelligenza Artificiale in classe*. Trento: Erickson.

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

Floridi, L. & Cabitza, F. (2021), *Intelligenza artificiale. L'uso delle nuove machine*. Milano: Bompiani.

Galletti, M. & Zipoli Caiani, S. (Eds.). (2024), *Filosofia dell'Intelligenza Artificiale. Sfide etiche e teoriche*. Bologna: Il Mulino.

Gardner, H. (2007), *Formae mentis. Saggio sulla pluralità dell'intelligenza* (13^a ed.). Milano: Feltrinelli.

Jaimez, A. (2023), *L'Intelligenza Emotiva nell'Era dell'Intelligenza Artificiale: come mantenere le nostre Capacità Umane in un Mondo Dominato dalla Tecnologia*. Independently published.

Jonassen, D.H., Carr, C. & Yueh, H.P. (1998), Computers as Mindtools for engaging learners in critical thinking. *Techtrends*, 43, 24-32. <https://doi.org/10.1007/BF02818172>.

Lorenzoni, G. (2024), *L'Intelligenza Artificiale a scuola. Guida per una pratica didattica consapevole*. Torino: Lattes.

Messina, S., Gaggioli, C. & Panciroli, C. (2024), Apprendere ed insegnare nell'era degli ecosistemi digitali intelligenti: pratiche didattiche e nuove piste di ricerca. *Media Education. Studi, ricerche e buone pratiche*, 15(1), 81-90. <https://doi.org/10.36253/me-15835>.

Miceli, E. (2023), *L'intelligenza artificiale con ChatGPT e ricadute nella didattica: da Don Milani all'IA. Approcci didattici veicolati dall'Intelligenza Artificiale per arrivare all'inclusione scolastica*. Independently published.

Mishra, P. & Koehler, M.J. (2006), Technological pedagogical content knowledge: a framework for teacher knowledge. *Teachers College Record*, 108(6), 1017-1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>.

Moriggi, S. & Pireddu, M. (2024). *L'Intelligenza Artificiale e i suoi fantasmi. Vivere e pensare con le reti generative*. Trento: Il Margine.

Nalbant, K.G. (2021), The Importance of Artificial Intelligence in Education: a short review. *Journal of Review in Science and Engineering*.

OECD (2024), Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449.

Panciroli, C. & Rivoltella, P.C. (2023), *Pedagogia algoritmica. Per una riflessione educativa sull'Intelligenza Artificiale*. Brescia: Morcelliana - Schol .

Panciroli, C., Rivoltella, P.C., Gabbrielli, M. & Richter, O.Z. (2020), Intelligenza artificiale e educazione: nuove prospettive di ricerca.

Form@re - Open Journal per la formazione in rete, 20(3), 1-12.
<https://doi.org/10.13128/form-10210>.

Pastorelli, V. (2024), L'Intelligenza Artificiale (IA) nell'educazione: Sfide, opportunità e considerazioni etiche. *Mizar. Costellazione di pensieri*, 21, 115-127. DOI: 10.1285/i24995835v2024n21p115.
<http://siba-ese.unisalento.it/index.php/mizar/article/viewFile/28702/23835>.

Pierdicca, R. & Frontoni, E. (2023), *Educare con le nuove tecnologie. La realtà estesa per l'apprendimento*. Maggioli Editore.

Ranieri, M. (2022). Le competenze digitali degli insegnanti. In Biagioli, R. & Oliviero, S. (Eds.). (2022). *Il Tirocinio Diretto Digitale Integrato (TDDI). Il progetto sperimentale per lo sviluppo delle competenze delle maestre e dei maestri*. Firenze: Firenze University Press, 49-60. DOI: 10.36253/978-88-5518-587-5.

Ranieri, M., Cuomo, S. & Biagini, G. (2023), *Scuola e intelligenza artificiale. Percorsi di alfabetizzazione critica*. Roma: Carocci.

Recalcati M. (2014), *L'ora di lezione*. Torino: Einaudi.

Russell, S. & Norvig, P. (2021), *Intelligenza artificiale. Un approccio moderno*. Pearson Editore.

Sancassani, S. & Mariconda, C. (2023), Smart Learning Design 25: Il metodo in 10 passi. Programma Teaching4Learning@Unipd®. Università di Padova.

Tammaro, G. (2024), *A scuola con l'intelligenza artificiale: un assistente "intelligente" al servizio degli insegnanti*. Independently published.

Unione Europea (2024), AI Act. Regolamento (UE) 1689/2024 del Parlamento Europeo e del Consiglio del 12 luglio 2024 relativo all'intelligenza artificiale. Gazzetta Ufficiale dell'Unione Europea, L.295, 1-60.

2. APPROCCIO PLURALE ALL'INTELLIGENZA ARTIFICIALE: SFIDE ETICHE E FORMATIVE NELLE ISTITUZIONI SCOLASTICHE

di Sara Pellegrini, Riccardo Sebastiani, Patrizia Ninassi, Emanuela Lampis*

Abstract: La ricerca esplora le percezioni di docenti e personale amministrativo di un istituto comprensivo in Sardegna, evidenziando disparità nelle competenze digitali e l'urgenza di formazione mirata. Si evidenzia il potenziale dell'IA nel personalizzare l'apprendimento e ottimizzare i processi amministrativi, con attenzione ai rischi di privacy, bias e dipendenza tecnologica. Lo studio propone un approccio etico e integrato per favorire inclusività e sostenibilità nei contesti educativi.

Parole chiave: Umanocentrico, Bias, Governance, Sostenibilità, Etica.

Abstract: The study examines the perceptions of teachers and administrative staff in a comprehensive school in Sardinia, highlighting disparities in digital skills and the urgent need for targeted training. It emphasizes the potential of AI to personalize learning and optimize administrative processes, while addressing risks related to privacy, bias, and technological dependency. The research proposes an ethical and

* Sara Pellegrini, Dipartimento di Scienze Umane, Università degli Studi "Link", <https://orcid.org/0009-0008-3827-3916>, s.pellegrini@unilink.it. Riccardo Sebastiani, Dipartimento di Scienze Umane, Università degli Studi "Link", <https://orcid.org/0009-0002-1090-0734>, r.sebastiani@unilink.it. Patrizia Ninassi, Dipartimento di Scienze Umane, Università degli Studi "Link", MIM, p.ninassi@gmail.com. Emanuela Lampis, Dirigente Scolastico MIM: emanuela.lampis@unilink.it.

Accettato Dicembre 2024 - Pubblicato Aprile 2025

integrated approach to foster inclusivity and sustainability in educational contexts.

Keywords: Human-centric; Bias; Governance; Sustainability; Ethics.

Introduzione

L'intelligenza artificiale (IA), con la sua capacità di trasformare la comunicazione tra i sistemi micro, macro, meso, ed endo, rappresenta una delle innovazioni più rivoluzionarie del nostro tempo. Tuttavia, il rapido sviluppo di tecnologie come l'apprendimento automatico e le affordance (de Kerckhove, 2021) di nuova generazione solleva questioni critiche riguardo ai rischi e alle implicazioni etiche, sociali ed economiche, nonché sulla loro reale applicabilità. Questo lavoro esplora l'approccio formativo all'IA, focalizzandosi sulla necessità di integrare la pluralità delle persone che sono coinvolte nel mondo scolastico per promuovere il benessere delle comunità attraverso un uso responsabile e sostenibile dell'IA. L'adozione dell'IA all'interno di un istituto comprensivo della Sardegna è una scelta complessa che richiede la formazione diffusa per prevenire i bias e le questioni della chiarezza del processo. Occorre pertanto porsi interrogativi fondativi relativi all'impiego dell'IA e delle *chatbot* per gli uffici di segreteria. Ciò va pensato in vista della effettiva bontà degli algoritmi avanzati perché possano comportare la reale riduzione delle ore lavorative prevista in numerose mansioni, ponendo importanti interrogativi sulla fiducia e sull'efficacia di queste tecnologie, evidenziando l'urgenza di un framework informativo robusto e di strategie educative efficaci per mitigare i rischi associati.

Attraverso questa ricerca, basata sulla matrice dei dati, si vogliono indagare le credenze dei docenti e degli impiegati rispetto a:

- gli sviluppi a livello pedagogico e educativo dell'IA che esplorano le percezioni sociali e culturali, valutando come queste influenzano le politiche pubbliche e la regolamentazione della didattica;
- le implicazioni etiche dell'IA quali i bias, la privacy e l'autonomia degli individui;
- le connessioni tra IA e sostenibilità educativa, amministrativa e organizzativa proponendo un'ecologia integrale che consideri le interazioni tra IA, ambiente e territorio e sul benessere.

Ciò facendo, si esamina come l'IA possa essere utilizzata per migliorare l'esperienza umana, favorendo interazioni più efficaci tra studenti, docenti e amministrazione e valutando le trasformazioni nei processi di lavoro e nelle organizzazioni, esplorando le potenziali conseguenze di identità istituzionale, economiche e sociali dell'automazione. Le riflessioni della nostra ricerca sono molteplici. Da un lato, contribuiamo a una comprensione più profonda delle sfide poste dall'IA che è la rivoluzione delle attitudini, fornendo un quadro teorico per la sua analisi critica. Dall'altro, esploriamo i risvolti pratici per l'introduzione dell'IA, suggerendo scelte ambientali che promuovono un uso responsabile e sostenibile di queste tecnologie. Il nostro approccio interagente tra il consuetudinario e l'innovativo, intende garantire che l'IA non solo rispetti i valori umani fondamentali, ma contribuisca anche al benessere delle persone e delle comunità. Questa ricerca vuole offrire un contributo al dibattito globale sull'IA nell'Istituto comprensivo della Sardegna, credendo fermamente che solo attraverso un'analisi critica e una regolamentazione responsabile l'IA potrà realizzare il suo pieno potenziale, pertanto, l'indagine si propone di condurre

un'esplorazione contestuale per bilanciare l'uso della tecnologia con il mantenimento del rapporto umano, essenziale per lo sviluppo del pensiero critico e delle competenze sociali degli studenti.

L'IA sta rapidamente diventando una forza trainante nel settore dell'istruzione, promettendo di ridefinire tanto l'insegnamento quanto l'apprendimento. È bene ricordare che in Italia il dibattito sull'adozione dell'IA nelle scuole non riguarda solo le potenzialità, ma si intreccia profondamente con riflessioni sulle sue implicazioni etiche, sulla necessità di adattare il ruolo dell'insegnante e sulla tutela dei valori educativi. Questo dimostra che il cammino verso un'integrazione osmotica richiederà attenzione e dialogo.

L'indagine, oggetto di questo studio, è stata svolta tra gli insegnanti sardi e il personale della segreteria, raccogliendo testimonianze che hanno offerto uno spaccato vivido delle aspettative e dei timori legati all'introduzione di questi strumenti avanzati. Se da un lato l'IA è percepita come un alleato prezioso per migliorare l'efficienza nelle attività di segreteria e nella gestione amministrativa, automatizzando le pratiche burocratiche che spesso appesantiscono il lavoro del personale di segreteria, dall'altro suscita interrogativi e timori, soprattutto a livello etico e di privacy. Molti insegnanti vedono in essa un potenziale strumento di supporto per adattare i percorsi di apprendimento alle esigenze individuali degli studenti, specie in contesti di inclusione e supporto ai bisogni educativi speciali (BES e DSA). In quest'ottica, l'IA potrebbe facilitare la personalizzazione dei percorsi di apprendimento, permettendo agli insegnanti di rispondere in modo più mirato alle esigenze di ciascuno. Ciononostante, vi sono delle preoccupazioni etiche che, come già sottolineato, emergono con forza soprattutto in merito al rischio di un eccessivo affidamento alla tecnologia. Molti insegnanti temono

che l'IA possa ridurre la capacità di pensiero critico e creativo degli studenti, rischiando di creare una generazione eccessivamente dipendente dagli strumenti digitali per affrontare o risolvere problemi complessi. Le preoccupazioni etiche non si fermano qui: le questioni legate alla privacy, alla sicurezza dei dati e alla protezione dell'identità digitale degli studenti e del personale scolastico, rappresentano tematiche centrali, spesso citate dagli intervistati. L'IA è vista, quindi, come un coltello a doppio taglio, in cui i benefici pratici devono essere bilanciati con cautele e regolamentazioni che ne garantiscano un uso etico e consapevole. Un aspetto significativo che emerge dall'indagine è un livello di competenza degli insegnanti e degli assistenti amministrativi, rispetto all'utilizzo dell'IA, molto disomogeneo. Alcuni partecipanti, infatti, hanno dichiarato una conoscenza limitata o superficiale delle tecnologie intelligenti, mentre altri mostrano familiarità con termini come "machine learning" e "elaborazione del linguaggio naturale". Questa disomogeneità suggerisce la necessità di progetti di formazione specifica e continuativa, volti a dotare tutti gli insegnanti e gli assistenti amministrativi degli strumenti necessari per comprendere, utilizzare e integrare l'IA nelle proprie pratiche didattiche e amministrative, riconoscendo il ruolo essenziale che entrambe le figure svolgono all'interno dell'istituzione educativa. Tale percorso formativo dovrebbe includere i docenti e il personale amministrativo per fornire a tutti i membri del personale scolastico le competenze per integrare l'IA nella didattica e nelle attività gestionali, promuovendo così una crescita integrata e consapevole dell'intera comunità educativa. Non si tratta semplicemente di acquisire competenze tecniche per l'uso di nuove tecnologie, ma anche di sviluppare una comprensione critica e consapevole di come questi strumenti possano influire su ruoli, compiti e responsabilità. Per i docenti, ciò significa esplorare come l'IA possa supportare la didattica, personalizzare i percorsi di apprendimento

e migliorare l'interazione con gli studenti. Per gli assistenti amministrativi, implica scoprire come l'IA possa facilitare la gestione dei dati, ottimizzare i processi amministrativi e migliorare l'efficienza organizzativa. In entrambi i casi, è fondamentale considerare non solo le competenze tecniche, ma anche quelle umane. Comprendere il potenziale impatto di queste tecnologie consente sia agli insegnanti che agli assistenti amministrativi di riflettere su come il proprio ruolo possa evolversi, sul valore aggiunto che l'IA può portare e sugli obiettivi educativi e formativi da raggiungere. Solo attraverso una formazione integrata e critica, docenti e personale amministrativo potranno essere protagonisti di un cambiamento consapevole e benefico per l'intera comunità scolastica. Gli insegnanti riconoscono che, per valorizzare appieno il potenziale di queste tecnologie, sarà necessario un approccio equilibrato che preservi l'importanza insostituibile del rapporto umano nella formazione dei giovani. Dai risultati di questa indagine emerge chiaramente che l'integrazione dell'IA nel sistema educativo sardo richiederà non solo investimenti nelle infrastrutture tecnologiche, ma anche un sostegno concreto per gli insegnanti, attraverso percorsi di formazione mirati e spazi di confronto sui temi etici e pedagogici. Solo in questo modo sarà possibile costruire una scuola capace di accogliere l'innovazione senza perdere di vista la propria missione educativa, intrinsecamente legata alla crescita personale e allo sviluppo del pensiero critico di ogni studente. Lo scopo di questo studio è quello di evidenziare un ambiente educativo in Sardegna dove l'IA è stata accolta con interesse e con un sano scetticismo

1. Contesto e quadro teorico

L'introduzione dell'IA nel sistema educativo rappresenta un'evoluzione che va oltre la semplice innovazione tecnologica: si tratta di un cambiamento profondo, che investe le modalità di apprendimento e insegnamento, l'organizzazione scolastica e il ruolo stesso degli insegnanti. Questo passaggio si colloca in un più ampio processo di digitalizzazione che sta coinvolgendo le scuole a livello globale e richiede un ripensamento delle competenze necessarie per operare efficacemente in un contesto sempre più interconnesso e tecnologicamente avanzato (Schwab & Davis, 2021). Il quadro teorico di riferimento per questa transizione è rappresentato da modelli come il TPACK (Technological Pedagogical Content Knowledge) e il DigCompEdu (Digital Competence Framework for Educators), che non solo promuovono lo sviluppo di abilità tecnologiche, ma forniscono una cornice interpretativa per comprendere come la tecnologia possa diventare un valore aggiunto in ambito educativo (Koehler & Mishra, 2006; Redecker, 2017).

Il modello TPACK, sviluppato per mettere in luce l'interdipendenza tra conoscenze tecnologiche, pedagogiche e disciplinari, sostiene che una reale competenza tecnologica in campo educativo non si limiti alla capacità di usare strumenti innovativi, ma richieda una capacità di integrare tali strumenti in modo sinergico con il contenuto disciplinare e le metodologie di insegnamento. Questo significa che l'introduzione dell'IA in classe non può prescindere da un'adeguata preparazione del docente che sappia adattare le sue conoscenze pedagogiche e disciplinari alle nuove potenzialità offerte dalla tecnologia (Mishra & Koehler, 2021). In pratica, l'IA dovrebbe supportare e arricchire il lavoro dell'insegnante, piuttosto che sostituirlo, contribuendo a una didattica che sia al contempo più efficace e inclusiva. Accanto al

TPACK, il DigCompEdu rappresenta un ulteriore riferimento teorico per comprendere l'importanza di sviluppare competenze digitali solide e responsabili negli insegnanti (Redecker, 2017). Questo framework europeo enfatizza il concetto di cittadinanza digitale, in cui la tecnologia diventa strumento per formare individui consapevoli, capaci di interagire con il digitale in modo critico e etico (Scherer *et al.*, 2021). L'adozione dell'IA nelle scuole, secondo questa prospettiva, non si limita ad ampliare l'offerta didattica, ma ha un valore educativo in sé, poiché contribuisce alla formazione di una generazione che sappia utilizzare i dati, gli algoritmi e le informazioni digitali con discernimento e responsabilità. In questo contesto teorico, l'indagine sugli insegnanti sardi offre uno spaccato utile per comprendere il livello di consapevolezza e preparazione dei docenti rispetto ai concetti chiave dell'IA e il loro atteggiamento nei confronti della sua integrazione nelle attività scolastiche. Le risposte raccolte evidenziano una disparità nelle conoscenze: mentre alcuni insegnanti dimostrano familiarità con le nozioni di machine learning, elaborazione del linguaggio naturale e visione artificiale, altri riconoscono una preparazione più limitata e desiderano maggiori strumenti di formazione per affrontare l'IA con sicurezza e competenza. Questa disparità riflette le diverse esperienze e percorsi formativi degli insegnanti, ma segnala anche l'urgenza di una formazione sistematica e continua che non solo spieghi il funzionamento della tecnologia, ma ne approfondisca le implicazioni etiche e pedagogiche (Selwyn, 2021). La necessità di un approccio equilibrato emerge chiaramente: l'IA, infatti, non deve diventare un fine, bensì un mezzo per arricchire la didattica, garantendo che gli studenti sviluppino le proprie capacità cognitive e sociali senza fare eccessivo affidamento sulla tecnologia. Un'integrazione efficace dell'IA richiede una visione ampia e lungimirante, che sappia valorizzare l'interazione tra l'insegnante e

lo studente e che ponga l'IA al servizio della relazione educativa, piuttosto che sostituirla (Williamson & Eynon, 2021). In questo senso, la prospettiva fornita dai modelli TPACK e DigCompEdu suggerisce che la tecnologia debba sempre essere integrata in modo critico e consapevole, accompagnata da riflessioni pedagogiche che ne amplificano l'impatto positivo e ne limitino i rischi. La voluta introduzione dell'IA nelle scuole sarde evidenzia l'importanza di un supporto formativo globale, sistemico ed ecologico, affinché tutti gli operatori scolastici possano dinamizzare nel flusso della quotidianità l'IA. Mentre gli insegnanti possono sviluppare competenze per integrare l'IA nella didattica e nella gestione dell'apprendimento, il personale amministrativo può acquisire strumenti utili per migliorare l'efficienza e la qualità dei servizi scolastici, approfondendo al contempo una consapevolezza critica rispetto ai limiti e alle implicazioni etiche di questa tecnologia. L'integrazione dell'IA, dunque, non dovrebbe ridursi a una semplice applicazione tecnica, ma inserirsi in una visione educativa complessiva che promuova competenze digitali e capacità critiche, sia negli insegnanti e negli studenti sia nel personale amministrativo. In questo modo, l'intero sistema educativo potrà preparare le nuove generazioni a vivere in un contesto in cui la tecnologia avrà un ruolo sempre più centrale e influente nella vita quotidiana, garantendo al contempo un funzionamento interno più moderno ed efficiente.

2. Metodologia

La metodologia utilizzata per questa indagine esplorativa è stata strutturata con l'obiettivo di raccogliere dati approfonditi sulle percezioni, le competenze e le aspettative degli insegnanti sardi rispetto all'introduzione dell'IA nel sistema educativo. Per

garantire una raccolta dati ampia e rappresentativa, è stato scelto di somministrare un questionario a un campione di insegnanti di vari ordini e gradi scolastici, distribuiti su tutto il territorio regionale. Il questionario è stato progettato per esplorare diversi aspetti dell'IA in ambito educativo, suddividendo le domande in sezioni tematiche quali: familiarità e livello di conoscenza dell'IA, percezione dei benefici e dei rischi, possibili applicazioni in ambito didattico e amministrativo, e considerazioni etiche sull'uso di queste tecnologie. Nella fase di progettazione del questionario, è stata posta particolare attenzione alla formulazione delle domande, in modo che fossero facilmente comprensibili per tutti i partecipanti, indipendentemente dal loro livello di competenza tecnologica. A tal fine, sono state incluse sia domande a risposta chiusa – per permettere un'analisi quantitativa dei dati – sia domande aperte, che hanno dato agli insegnanti la possibilità di esprimere opinioni personali, esperienze dirette e suggerimenti specifici. Le domande a risposta chiusa comprendevano opzioni predefinite, che spaziavano dalle opinioni sull'impatto dell'IA sulla qualità dell'insegnamento, alle preferenze riguardo le aree in cui l'IA potrebbe supportare l'attività didattica o ridurre il carico burocratico. Le domande aperte, invece, hanno permesso di ottenere risposte più articolate su temi quali le preoccupazioni etiche, le condizioni necessarie per un'efficace implementazione dell'IA nelle scuole e le esigenze di formazione percepite dai docenti. L'indagine è stata condotta su un arco temporale di alcune settimane, durante le quali i questionari sono stati distribuiti sia in formato cartaceo sia digitale, in modo da raggiungere un numero maggiore di insegnanti e agevolare la partecipazione, soprattutto nelle aree con una più limitata accessibilità alle risorse tecnologiche. La somministrazione digitale ha garantito una raccolta dati più rapida e un'archiviazione centralizzata delle risposte, facilitando l'analisi successiva. La metodologia mista di

raccolta, inoltre, ha permesso di includere un campione rappresentativo, come descritto nella metodologia, che ha autonomamente scelto di partecipare ha generato una ampia e diversificata stratificazione, comprendendo insegnanti di diverse fasce d'età, con vari livelli di esperienza e competenza digitale.

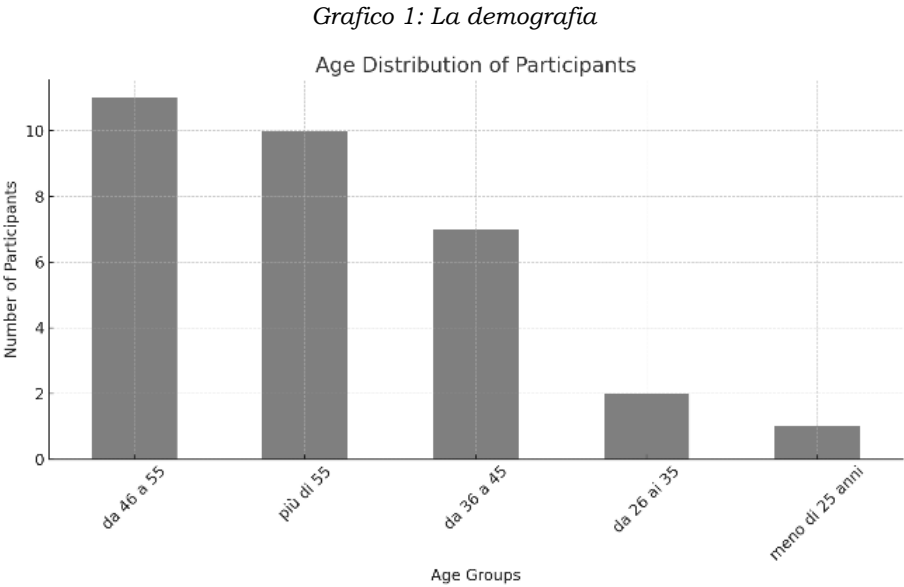
Per quanto riguarda l'analisi dei dati, le risposte a scelta multipla sono state elaborate utilizzando tecniche di analisi statistica descrittiva, al fine di identificare tendenze e ricorrenze tra le opinioni degli insegnanti. I dati quantitativi hanno offerto una visione d'insieme delle percezioni generali riguardo all'IA, consentendo di delineare i temi più sentiti dai partecipanti. Le risposte aperte sono state analizzate qualitativamente utilizzando una procedura di codifica per individuare i principali temi emergenti e categorizzare le risposte in base a criteri di frequenza e rilevanza. Questa combinazione di approcci quantitativi e qualitativi ha permesso di ottenere un quadro ricco e dettagliato, non solo delle opinioni prevalenti ma anche delle sfumature, dei dubbi e delle riflessioni personali degli insegnanti. Nel complesso, questa metodologia ha fornito una base solida per comprendere le percezioni e le esigenze degli insegnanti e degli assistenti amministrativi sardi rispetto all'introduzione dell'IA nelle scuole, evidenziando non solo le aspettative e le resistenze, ma anche gli ambiti in cui sarebbe necessaria una formazione specifica per un'adozione consapevole e responsabile delle tecnologie intelligenti in ambito educativo.

3. Risultati

Il questionario è stato somministrato a un campione di insegnanti e di assistenti amministrativi della regione Sardegna, che operano in diversi gradi scolastici e istituti scolastici. Questo gruppo

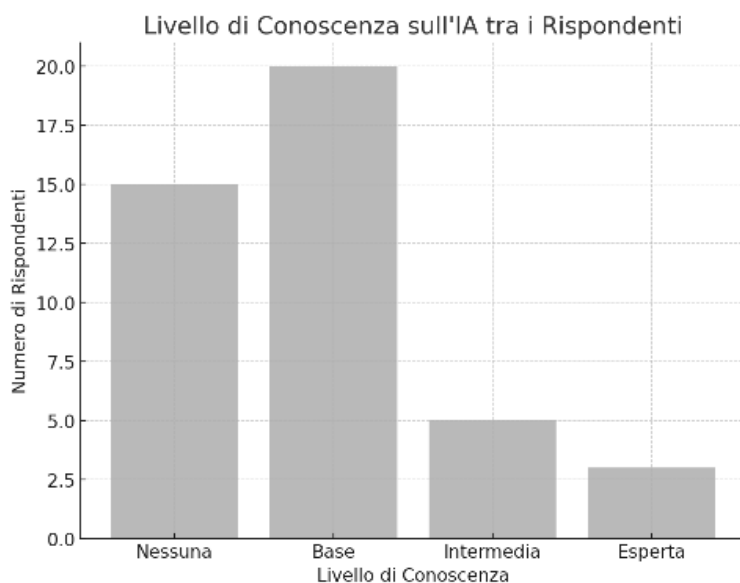
eterogeneo è stato scelto per ottenere una rappresentazione ampia e diversificata delle percezioni e delle esperienze dei docenti e degli assistenti amministrativi, riguardo all'integrazione dell'IA nell'istruzione. L'indagine ha coinvolto insegnanti e assistenti amministrativi con vari livelli di esperienza e competenza tecnologica, allo scopo di raccogliere dati dettagliati e rappresentativi sulle aspettative, gli interessi e le preoccupazioni etiche legate all'uso dell'IA nelle scuole sarde. L'analisi dei risultati dell'indagine sull'introduzione dell'IA nell'istruzione sarda evidenzia un interesse crescente e una percezione variegata da parte dei docenti e degli assistenti amministrativi rispetto alle potenzialità, alle sfide e alle implicazioni etiche dell'IA nel contesto educativo. Il campione di insegnanti e assistenti amministrativi coinvolto, è stato selezionato un campione rappresentativo di 227 partecipanti, tra i 27 e i 57 anni d'età, dell'Istituto comprensivo della Sardegna, utilizzando un campionamento stratificato per area geografica e tipo di scuola. La rappresentatività è stata garantita attraverso la proporzionalità tra i sottogruppi e la popolazione complessiva di riferimento, con un margine di errore del 5%. Ciò è rappresentativo[1] di diversi livelli scolastici e distribuito su tutto il territorio regionale, ha espresso opinioni ricche e dettagliate in merito a vari aspetti dell'IA, mostrando in molti casi un'apertura all'integrazione tecnologica pur accompagnata da un'attenta riflessione sui rischi e sulle responsabilità legate all'uso di queste tecnologie. Il grafico (Grafico 1) sulla distribuzione anagrafica dei partecipanti all'indagine, provenienti dai comuni di Burcei e di Maracalagonis, rivela interessanti spunti. Burcei, situato nella parte centro-orientale della provincia del Sud Sardegna, sorge sui monti del Sarrabus a quasi 700 metri sul livello del mare e conta 2.646 abitanti (ISTAT, 1° gennaio 2022). La posizione isolata e la bassa densità di popolazione potrebbero aver influenzato una partecipazione prevalentemente matura, con una forte presenza di individui di età compresa tra i 36

e i 55 anni. Dall'altro lato, Maracalagonis, situato nella parte sudorientale del Campidano a soli 17 km dal capoluogo, Cagliari, è uno dei 17 comuni della Città Metropolitana. Questo paese, che conta 7.873 abitanti (ISTAT, 1° gennaio 2022), ha visto una crescita demografica negli ultimi anni grazie a un mercato abitativo favorevole e un'economia sempre più fondata sul terziario. Anche qui, i partecipanti appartengono principalmente a fasce di età più mature. La comunità straniera a Maracalagonis, sebbene ridotta (1,95% della popolazione), include un 34% di persone di origine marocchina che apportano diversità culturale e sociale. Questa prevalenza di partecipanti maturi e con esperienza potrebbe riflettere una maggiore familiarità e consapevolezza rispetto ai temi dell'educazione e dell'IA. L'assenza di giovani educatori sottolinea una potenziale mancanza di interesse verso l'IA in ambito educativo da parte delle fasce più giovani, forse meno coinvolte nei dibattiti sull'innovazione educativa.



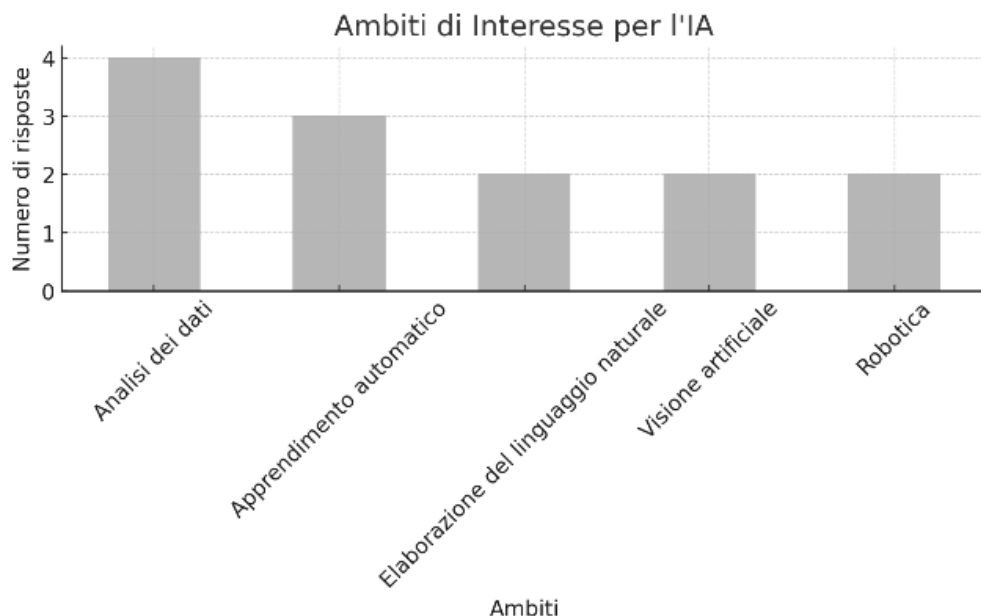
La maggior parte dei rispondenti possiede una conoscenza “base” o “intermedia” dell’IA, con pochi docenti che si considerano esperti (Grafico 2). Questa disparità di competenze mette in evidenza una necessità condivisa di formazione specifica, che permetta ai docenti di comprendere appieno le potenzialità dell’IA e di utilizzare tali strumenti in modo efficace e consapevole.

Grafico 2: Docenti e IA



L’interesse per l’IA si concentra principalmente su ambiti come l’apprendimento automatico, l’analisi dei dati e l’elaborazione del linguaggio naturale, applicazioni ritenute particolarmente utili per personalizzare l’insegnamento e supportare la gestione delle informazioni didattiche (Grafico 3).

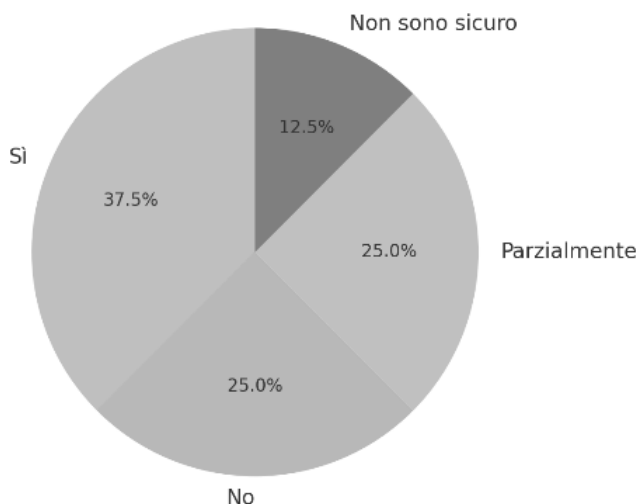
Grafico 3: Gli ambiti di interesse



Alla domanda su come l'IA possa migliorare la qualità dell'istruzione, la maggior parte dei rispondenti ha risposto positivamente, anche se con qualche riserva. Mentre una buona parte è convinta che l'IA possa migliorare l'efficacia dell'insegnamento e le pratiche amministrative, alcuni esprimono dubbi e ritengono importante bilanciare l'uso della tecnologia con l'interazione umana (Grafico 4).

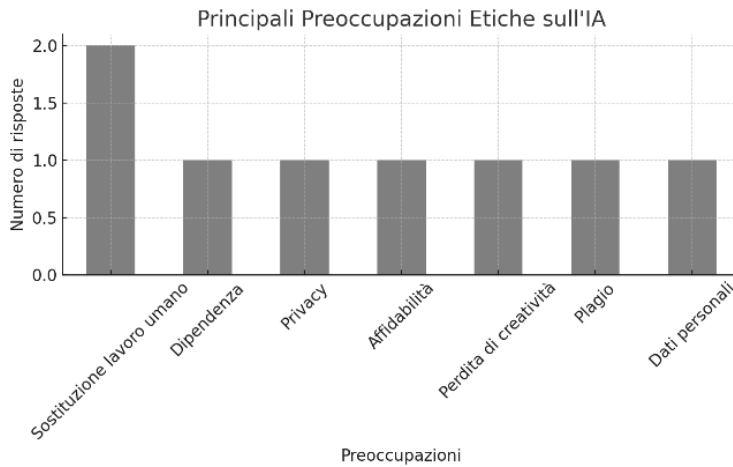
Grafico 4: Riflessioni sul miglioramento con IA

Percezione di Miglioramento della Qualità dell'Istruzione



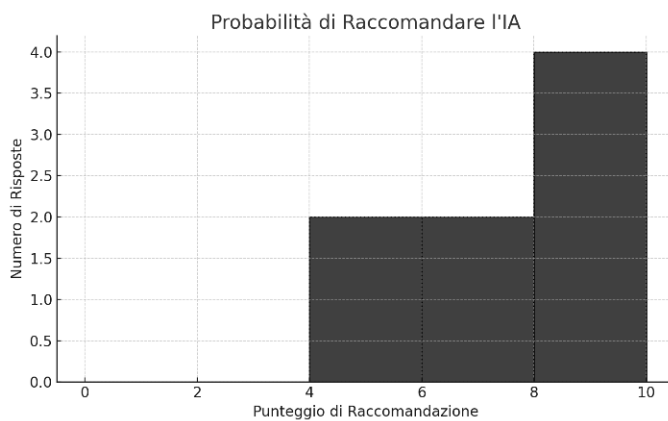
Docenti e assistenti amministrativi hanno espresso serie preoccupazioni etiche sull'uso dell'IA a scuola. Tra i temi più sentiti emergono i rischi per la privacy degli studenti, la dipendenza dalla tecnologia, la potenziale perdita di creatività e pensiero critico, oltre al timore che l'IA riduca progressivamente il ruolo umano, limitando l'interazione diretta tra insegnante e studente, e tra personale amministrativo e pubblico. Questi aspetti evidenziano una visione critica e consapevole sull'IA, ritenendo fondamentale un uso regolamentato e bilanciato, che la renda strumento di supporto e non di sostituzione dell'intervento umano (Grafico 5).

Grafico 5. Preoccupazioni, IA ed etica



I punteggi relativi alla probabilità di promuovere l'uso dell'IA variano significativamente (Grafico 6). Alcuni insegnanti mostrano un atteggiamento positivo verso la diffusione dell'IA, mentre altri sono più cauti, con un interesse moderato bilanciato dalla consapevolezza dei suoi limiti e rischi.

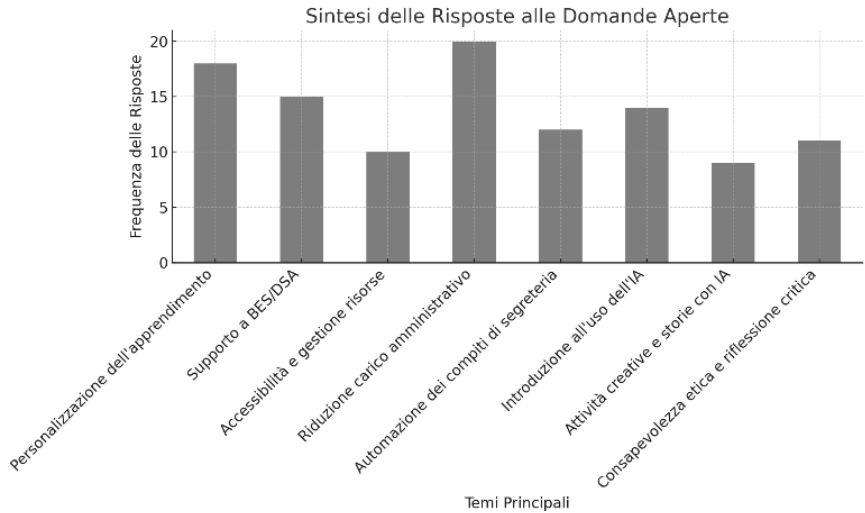
Grafico 6: Uso di IA



Per quanto riguarda le domande aperte i docenti hanno indicato vari modi in cui l'IA potrebbe apportare benefici alla scuola. Molti hanno sottolineato la capacità dell'IA di personalizzare l'apprendimento, adattando i contenuti e i metodi didattici alle esigenze individuali degli studenti. In particolare, è stato evidenziato il potenziale dell'IA per supportare gli studenti con bisogni educativi speciali (BES) o disturbi specifici dell'apprendimento (DSA), fornendo strumenti che possano agevolarne l'inclusione e l'apprendimento autonomo. Altri hanno suggerito che l'IA potrebbe migliorare l'accessibilità ai materiali didattici, rendendo più semplice la gestione delle risorse scolastiche e facilitando la comunicazione tra insegnanti, studenti e famiglie. Molti docenti vedono nell'IA un valido supporto per ridurre il carico amministrativo e migliorare l'efficienza gestionale. Tra i compiti più frequentemente citati vi sono la gestione dei registri scolastici, l'organizzazione degli orari e la programmazione delle attività didattiche. Alcuni hanno proposto l'uso di chatbot basati su IA per rispondere alle domande delle famiglie in tempo reale, facilitando la comunicazione con la segreteria e risparmiando tempo al personale amministrativo. La possibilità di automatizzare questi processi è vista come un modo per lasciare agli insegnanti più tempo da dedicare all'insegnamento diretto e all'interazione con gli studenti. Gli insegnanti hanno suggerito diverse attività per introdurre gli studenti all'uso consapevole dell'IA. Alcuni propongono esercitazioni pratiche che prevedano l'utilizzo di strumenti di IA per ricerche o esperimenti di base, consentendo agli studenti di comprendere il funzionamento della tecnologia e di acquisire competenze digitali. Altri suggeriscono attività creative, come la realizzazione di storie o progetti in cui l'IA possa essere utilizzata per esplorare culture e periodi storici in modo interattivo. È stato anche proposto di incoraggiare gli studenti a riflettere criticamente sugli aspetti etici dell'IA, aiutandoli a sviluppare una

consapevolezza riguardo alle potenziali implicazioni sociali e culturali dell'uso di queste tecnologie.

Grafico 7: Considerazioni



Questi riassunti evidenziano un forte interesse da parte dei docenti per un approccio educativo che renda l'IA uno strumento inclusivo e di supporto, non solo per la didattica, ma anche per migliorare l'efficienza scolastica e promuovere la consapevolezza tecnologica tra gli studenti (Grafico 7).

4. *Discussione*

La discussione sull'integrazione dell'IA nelle istituzioni scolastiche rivela un equilibrio complesso tra le potenzialità della tecnologia e le preoccupazioni etiche sollevate dai docenti sardi. I dati raccolti mostrano come l'IA sia percepita positivamente per il

potenziale di supportare attività di personalizzazione dell'apprendimento e di automatizzazione dei compiti amministrativi, riducendo il carico burocratico degli insegnanti e migliorando l'efficienza delle scuole. Tuttavia, come riportato da molti studi recenti (Bender *et al.*, 2021; Crawford & Paglen, 2019), i rischi connessi all'uso dell'IA, come la dipendenza tecnologica e la tutela della privacy, richiedono un approccio critico e regolamentato. Questi rischi sottolineano la necessità di strategie educative e normative solide, in linea con la letteratura internazionale che sottolinea l'importanza di un approccio umanocentrico e responsabile alla tecnologia (Binns, 2021).

I docenti esprimono in modo chiaro preoccupazioni legate alla possibilità che l'IA influenzi negativamente lo sviluppo delle capacità critiche e sociali degli studenti, trasformando l'ambiente scolastico in un contesto in cui il pensiero critico rischia di essere subordinato a risposte automatizzate. Secondo Floridi (2021), un uso non ponderato dell'IA nelle scuole potrebbe ridurre significativamente le interazioni tra studenti e insegnanti, un aspetto cruciale per il processo educativo e per lo sviluppo delle competenze sociali. La prospettiva dei docenti riflette questa preoccupazione, evidenziando il timore che l'IA, automatizzando determinate funzioni, possa progressivamente sminuire il ruolo dell'insegnante, trasformandolo da guida educativa a semplice supervisore delle tecnologie. Questo approccio rischia di minare la relazione insegnante-studente, che è alla base dell'apprendimento efficace, basato sulla fiducia, l'empatia e l'attenzione alle esigenze individuali. Accanto alle implicazioni relazionali, emerge una questione urgente riguardante la formazione dei docenti. Molti insegnanti dichiarano di non avere una conoscenza approfondita delle tecnologie IA, limitandosi a una comprensione superficiale di concetti come "machine learning" e "natural language processing" (elaborazione del linguaggio naturale). La letteratura evidenzia che

questa mancanza di competenze rappresenta un rischio significativo per l'introduzione efficace e responsabile dell'IA nelle scuole. Binns (2021) sostiene che una formazione continua e mirata è essenziale per evitare un uso non critico della tecnologia. L'introduzione dell'IA nell'ambiente scolastico richiede una preparazione specifica che permetta ai docenti di comprendere le potenzialità e i limiti di questi strumenti. Senza un'educazione tecnologica adeguata, l'IA rischia di essere percepita come una soluzione automatica e indiscutibile, piuttosto che come un supporto consapevole al processo educativo. Questa percezione potrebbe limitare l'autonomia degli insegnanti, lasciando alla tecnologia il ruolo decisionale principale, un rischio che la letteratura evidenzia come particolarmente pericoloso nei contesti educativi. La questione delle disuguaglianze nell'accesso alle risorse tecnologiche si rivela altrettanto centrale. Le disparità tra le scuole, soprattutto in Sardegna, riguardano la disponibilità di infrastrutture adeguate e di risorse tecnologiche avanzate, fondamentali per l'implementazione dell'IA in modo equo ed efficace. Molti istituti non dispongono delle attrezzature e delle competenze necessarie per sfruttare appieno le potenzialità dell'IA, un divario che rischia di amplificare le disuguaglianze già presenti tra studenti provenienti da contesti socio-economici differenti. La letteratura sottolinea che l'introduzione delle tecnologie digitali deve essere accompagnata da politiche che promuovano la sostenibilità e l'inclusività, con l'obiettivo di colmare le disparità piuttosto che accentuarle (Floridi, 2021). L'IA, quindi, non dovrebbe essere considerata esclusivamente uno strumento per l'innovazione, ma anche un mezzo per promuovere un ambiente educativo più equo, che valorizzi il potenziale di ogni studente indipendentemente dalle condizioni socio-economiche di partenza. I risultati dell'indagine suggeriscono, quindi, che un approccio equilibrato, critico e ben regolamentato sia indispensabile per

garantire che l'IA si integri nell'ambiente scolastico rispettando e valorizzando i principi educativi fondamentali. Questa tecnologia ha il potenziale di migliorare molti aspetti del sistema educativo, ma soltanto se utilizzata in modo consapevole, con una chiara consapevolezza dei rischi e una ferma attenzione al benessere degli studenti e alla protezione della privacy.

Conclusioni

La riflessione sull'integrazione dell'IA nelle scuole sarde evidenzia non solo le opportunità, ma anche le profonde sfide etiche e formative che derivano da un uso consapevole di questa tecnologia. L'indagine ha mostrato come l'IA possa essere percepita come un potente alleato per personalizzare l'insegnamento, ridurre i carichi amministrativi ottimizzando le pratiche che poi si riverberano sul sistema scolastico e migliorare l'accessibilità dell'educazione, in particolare per studenti con bisogni educativi speciali. Tuttavia, l'adozione dell'IA non può limitarsi a un approccio tecnico quale mero superamento del termico per l'avvento dell'analogico, ma deve necessariamente integrare una visione etica e umanocentrica. L'intero sistema educativo della scuola sarda, attraverso le risposte, manifesta una chiara consapevolezza dei rischi associati all'automazione educativa e alla dipendenza tecnologica, sottolineando l'importanza di mantenere l'interazione umana al centro del processo educativo. Questo equilibrio rispecchia una sfida comune in molti sistemi educativi internazionali (Floridi, 2021). La necessità di una formazione continua per i docenti e per gli assistenti tecnici e amministrativi emerge come un requisito imprescindibile. La disparità nei livelli di competenza tecnologica tra gli insegnanti e gli operatori di questo istituto comprensivo della Sardegna, evidenziata dai dati raccolti,

sottolinea l'urgenza di percorsi formativi che vadano oltre la semplice alfabetizzazione digitale, promuovendo una comprensione critica dell'IA e delle sue implicazioni educative e sociali (Binns, 2021). Questo aspetto è fondamentale per evitare un'adozione passiva della tecnologia, dove l'IA viene accettata acriticamente come soluzione universale, senza una riflessione approfondita sulle sue limitazioni e sui suoi rischi. La capacità di utilizzare l'IA in modo consapevole e critico non solo rafforza l'efficacia educativa, ma garantisce anche che la tecnologia sia al servizio dell'apprendimento e non viceversa. La questione delle disparità di accesso alle risorse tecnologiche è particolarmente rilevante in Sardegna, dove le scuole spesso operano con risorse limitate e infrastrutture inadeguate. Un'integrazione efficace dell'IA, quindi, richiede non solo la formazione del personale, ma anche investimenti strutturali per colmare le lacune tecnologiche tra i diversi istituti e regioni. Un'implementazione disomogenea dell'IA rischia infatti di amplificare le disuguaglianze educative, negando a una parte degli studenti le opportunità offerte dalle tecnologie avanzate (Bender *et al.*, 2021). Credendo che l'introduzione dell'IA nelle istituzioni scolastiche sarde rappresenta una strada percorribile solo attraverso un equilibrio tra innovazione e valori educativi fondamentali, un approccio pluralista e inclusivo, come suggerito dai modelli teorici di riferimento TPACK e DigCompEdu, può contribuire ad organizzare una profilazione di pratiche e di azioni con l'IA, pedagogicamente orientato, volto alla generazione di un sistema ecologico che valorizza osmoticamente la formazione integrale degli studenti e dei lavoratori, curando e preservando il ruolo umano nella co-costruzione della conoscenza. L'esplorazione porta alla luce una disomogeneità significativa nelle competenze digitali dei partecipanti, limitando una visione uniforme sull'adozione dell'IA. Non solo, le preoccupazioni etiche e la scarsità di infrastrutture tecnologiche potrebbero ostacolare una

piena implementazione. Il dato non è allarmante poiché si tratta di una istituzione che sta osservando scientificamente sé stessa per poter costruire azioni robuste e pedagogicamente fondate. Il progetto mostra un pensiero innovativo e si distingue per un approccio critico e umanocentrico che mira a bilanciare innovazione e valori educativi, promuovendo inclusività e sostenibilità. Si raccomanda di sviluppare programmi formativi mirati, politiche per ridurre le disuguaglianze digitali e ulteriori studi per valutare l'impatto a lungo termine dell'IA nel contesto scolastico. Lo studio esplora le opportunità e le sfide dell'intelligenza artificiale di un sistema scolastico, evidenziando come la tecnologia trasformi le dinamiche cognitive e sociali. Sebbene l'IA offra strumenti promettenti per personalizzare l'apprendimento e ottimizzare i processi amministrativi, dobbiamo già da ora pensare alla possibile emersione di criticità legate alla crescente tendenza all'"offloading", ossia l'esternalizzazione delle funzioni cognitive verso dispositivi digitali. Tale fenomeno, già analizzato da studiosi Stiegler (Turner, 2019), implica una delega progressiva delle capacità mnemoniche, analitiche e decisionali alle macchine, con il rischio di un indebolimento del capitale cognitivo personale. Questa tendenza, combinata con una riduzione della lettura attiva e dell'immagazzinamento mentale delle informazioni, produce una "porosità cognitiva", in cui la nostra memoria diventa sempre più dipendente da supporti esterni. Ciò genera preoccupazioni etiche e pratiche, non solo per la perdita di competenze fondamentali, ma anche per la riduzione dello spazio mentale necessario per il pensiero critico e l'elaborazione autonoma delle decisioni. Questa erosione del capitale cognitivo personale si riflette anche nelle scuole, dove il ruolo degli educatori rischia di essere limitato a supervisor tecnologici piuttosto che a promotori del pensiero critico. Oltre a tali preoccupazioni che debbono essere accolte e processate per attivare buone pratiche

della cura della cura (Mortari, 2022), possiamo riflettere su una serie di vantaggi a cui facciamo fatica a rinunciare. Concetti emergenti come il “senso comune digitale” (Floridi, 2021) offrono una prospettiva costruttiva, implicando la costruzione di un corpus condiviso di conoscenze pubbliche e accessibili, regolato da modelli di intelligenza artificiale che filtrano contenuti dannosi o polarizzanti: l’equilibrio tra il dominio pubblico e il diritto d’autore richiede una regolamentazione chiara per garantire un accesso equo alle risorse digitali.

Un’altra prospettiva interessante è il costrutto di “*crowdprompting*”, una metodologia che utilizza le intelligenze collettive per generare domande di qualità attraverso la collaborazione tra esperti e non esperti. Questo approccio potrebbe trasformare le pratiche educative, favorendo un’interazione più dinamica tra l’umano e il digitale. Similmente, l’idea di “*commonself*” esplora come le conoscenze condivise possano supportare l’autonomia decisionale e la crescita individuale in un contesto tecnologico. Non ultimo, l’avvento della GenAI e dei modelli di apprendimento avanzato solleva questioni sul futuro delle competenze umane. Se, da un lato, tecnologie come Neuralink potrebbero amplificare le capacità cognitive individuali, dall’altro pongono interrogativi fondamentali sull’equilibrio tra autonomia umana e dipendenza tecnologica. In questo contesto, il sistema educativo ha la responsabilità di promuovere una formazione critica e umanocentrica, che valorizzi il rapporto umano e contrasti la perdita di competenze essenziali. Se l’IA rappresenta un potente alleato per l’educazione e l’innovazione, allora la sua integrazione dovrebbe avvenire attraverso un approccio etico e responsabile. Solo bilanciando l’accumulo collettivo di conoscenze con lo sviluppo del capitale cognitivo personale, sarà possibile evitare un’erosione delle capacità umane e garantire un futuro in cui l’IA

sia un valido strumento di supporto e non sostituzione del pensiero critico.

Bibliografia

Bawden, D., & Robinson, L. (2023). *Sustainable Information: Ethics, Technology, and Transformation*. Oxford University Press.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610-623. <https://doi.org/10.1145/3442188.3445922>.

Bennett, S., & Oliver, M. (2021). Exploring the Role of Technology in Educational Change: New Theories, New Evidence. *Journal of Educational Technology Research and Development*, 69(4), 1120–1135. <https://doi.org/10.1007/s11423-021-10083-9>.

Binns, R. (2021). Fairness in machine learning: Lessons from political philosophy. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 1-12. <https://doi.org/10.1145/3442188.3445939>.

Cosentino, G. (2020). *Social media and the post-truth world order*. London; Cham: Palgrave Pivot.

Crawford, K., & Paglen, T. (2019). Excavating AI: The Politics of Images in Machine Learning Training Sets. *Excavating AI*. Retrieved from <https://excavating.ai/>.

de Kerckhove, D. (2021). The personal digital twin, ethical considerations. *Philosophical Transactions of the Royal Society A*, 379(2207), 20200367. <https://doi.org/10.1098/rsta.2020.0367>.

Floridi, L. (2020). *The Logic of Information: A Theory of Philosophy as Conceptual Design*. Oxford: Oxford University Press.

Floridi, L. (2021). *The ethics of artificial intelligence*. Oxford: Oxford University Press.

Henz, P. (2022). The societal impact of the Metaverse. *Discover Artificial Intelligence*, 2(1), 19.

Jones, K. (2021). Artificial Intelligence in Education: Promise, Implications, and Caution. *AI & Society*, 36(2), 143-154. <https://doi.org/10.1007/s00146-020-01003-0>.

Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional Complexity and the Science to Constrain It. *Science*, 342(6161), 935-937. <https://doi.org/10.1126/science.1238056>.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>.

Mishra, P., & Koehler, M. J. (2021). Technological Pedagogical Content Knowledge (TPACK): A Framework for Integrating Technology in Education. *Educational Technology*, 61(4).

Mortari, L. (2023). *La cultura della cura: Sguardo, ascolto e responsabilità*. Bentivoglio: ASMepa.

Noble, S. U. (2021). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.

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>.

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishing Group.

Parasuraman, R., & Manzey, D. H. (2010). Complacency and Bias in Human Use of Automation: An Attentional Integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 52(3), 381-410. <https://doi.org/10.1177/0018720810376055>.

Redecker, C. (2017). European Framework for the Digital Competence of Educators: DigCompEdu. Publications Office of the European Union.

Scherer, R., Howard, S. K., Tondeur, J., & Siddiq, F. (2021). Profiling teachers' readiness for online teaching and learning in higher education: Who's ready and who's not? *Computers in Human Behavior*, 118, 106675.

Schwab, K., & Davis, N. (2021). *Shaping the Future of the Fourth Industrial Revolution*. New York: Crown Currency.

Selwyn, N. (2021). *Education and Technology: Key Issues and Debates*. London; New York; Dublin: Bloomsbury Publishing.

Turner, B. (2019). Politicising the epokhé: Bernard Stiegler and the politics of epochal suspension. In I. Apostolescu (Ed.), *The subject(s) of phenomenology: Rereading Husserl*, Dordrecht: Springer. 341–354.

Williamson, B., & Eynon, R. (2021). *The Automation of Education: Critical Perspectives on AI, Algorithms, and Big Data*.

3. L'INTELLIGENZA ARTIFICIALE NELL'EDUCAZIONE: UN'ANALISI DEGLI STUDENTI ITALIANI

di Antonio Opromolla*

Abstract: L'Intelligenza Artificiale è ormai presente nelle attività quotidiane, e il settore educativo ne è un esempio. L'impatto più consistente riguarda in particolare l'esperienza educativa dei giovani adulti, che sempre di più la integrano nelle loro pratiche di studio. Il contributo mostra i risultati preliminari di uno studio che ha previsto la somministrazione di una breve survey a cui hanno preso parte alcune studentesse universitarie, che ha avuto lo scopo di conoscere la percezione dell'impiego dei sistemi di Intelligenza Artificiale nel corso dei loro processi formativi. Tali risultati indicano che, sebbene i sistemi di Intelligenza Artificiale vengano quotidianamente impiegati da parte delle rispondenti, le quali in generale si pongono in maniera mediamente positiva nei confronti di tale argomento, le funzionalità impiegate sono ancora limitate, mentre dubbi e paure sono numerose.

Parole chiave: Intelligenza Artificiale, giovani adulti, educazione, human-centered design, media education.

Abstract: Artificial Intelligence is now present in everyday activities; the educational sector is an example. The most substantial impact concerns the academic experience of young adults, who increasingly integrate it into their study practices. The contribution shows the preliminary results of a study involving the administration of a short survey in which female university students participated, which aimed to determine their

* Dipartimento di Scienze Umane, Università degli Studi dell'Aquila, <https://orcid.org/0000-0002-6415-3829>, antonio.opromolla@univaq.it.

Accettato Dicembre 2024 - Pubblicato Aprile 2025.

perceptions of using Artificial Intelligence systems in their educational processes. These results indicate that although Artificial Intelligence systems are used daily by the respondents, who generally take an average positive attitude towards this topic, the functionalities used are still limited, while doubts and fears are numerous.

Keywords: Artificial Intelligence, young adults, education, human-centered design, media education.

1. L'Intelligenza Artificiale nel settore educativo

L'Intelligenza Artificiale è ormai integrata nelle attività quotidiane, e il settore educativo ne sta subendo un crescente impatto, soprattutto tra i giovani adulti. Secondo i risultati di uno studio precedentemente svolto (Opromolla, 2024), la letteratura recente sull'Intelligenza Artificiale in ambito educativo evidenzia i suoi benefici sia per insegnanti che per studenti: essa può migliorare la personalizzazione dell'apprendimento, accelerare valutazioni e gestione amministrativa, e supportare l'insegnamento adattativo. Le sue applicazioni includono tutor intelligenti, aule virtuali e strumenti di feedback automatico. La letteratura evidenzia, tuttavia, come esistano anche questioni etiche, come la privacy e l'equità e come sia essenziale bilanciare i benefici con un approccio trasparente e multidisciplinare per affrontare le sfide legate alla sicurezza e alla giustizia sociale nel contesto educativo.

2. Un'analisi degli studenti italiani

Lo studio qui presente mostra i risultati preliminari di una breve survey rivolta a studenti universitari, che ha lo scopo di conoscere la percezione dell'impiego dei sistemi di Intelligenza

Artificiale nel corso dei loro processi formativi. Tale survey, somministrata online, ha coinvolto circa 50 partecipanti, la maggior parte studentesse, di età compresa tra i 19 e i 24 anni, di un corso di laurea in Scienze dell'Educazione. In termini generali, le rispondenti ritengono di avere una media conoscenza delle tecnologie digitali e di farne un utilizzo contenuto. Risultati meno positivi emergono in relazione all'impiego dell'Intelligenza Artificiale: il gruppo di rispondenti dichiara, infatti, di avere una conoscenza medio-bassa di tale tecnologia. Tra i sistemi di Intelligenza Artificiale più conosciuti ci sono chatbot e assistenti virtuali, ma anche oggetti intelligenti, mentre quelli meno conosciuti sono sistemi di raccomandazioni (ad esempio basati su scelte precedenti svolte dall'utente) e sistemi predittivi. In merito all'applicazione dell'Intelligenza Artificiale nello studio, il 40% delle rispondenti dichiara di aver utilizzato tali strumenti, mentre il 60% dichiara di non averne mai fatto uso.

Tra i sistemi più comunemente usati ci sono sistemi di traduzione e sistemi di revisione e correzione automatica, mentre tra quelli meno usati ci sono tutor virtuali (es. chatbot educativi) e piattaforme di apprendimento personalizzato. La percezione di utilità dei vari strumenti risulta molto alta: tra coloro che hanno dichiarato di aver usato sistemi di Intelligenza Artificiale nello studio, circa il 70% li ritiene utili o molto utili. Tale utilità sembra essere funzionale a processi di studio che vengono svolti in maniera "tradizionale" e "attiva" da parte delle rispondenti (es.: segnalazione di piccoli errori grammaticali su esercizi di lingua inglese, segnalazione di fonti per approfondimenti di argomenti specifici, etc.). Nel chiedere alle rispondenti una descrizione breve e istintiva della propria percezione sull'applicazione dell'Intelligenza Artificiale nell'educazione, i commenti si dividono tra quelli più positivi e quelli più negativi. Tra i primi, le rispondenti evidenziano come tale tecnologia sia utile a valorizzare e ottimizzare il tempo,

ad esempio accedendo a materiali che aiutano con la ripetizione degli argomenti. Tra i commenti meno positivi, le rispondenti evidenziano la necessità di usare la tecnologia con molta attenzione e cautela, la mancanza di fiducia totale data dalla percezione di “non-umanità” dell’Intelligenza Artificiale e dal rischio che possa fornire informazioni non corrette, nonché la mancanza di un’adeguata conoscenza e consapevolezza da parte degli studenti sui modi più consoni di impiegarla (cosa che potrebbe invogliare verso comportamenti superficiali).

Nell’identificare, più nel dettaglio, i vantaggi dell’Intelligenza Artificiale nei processi educativi e di apprendimento, le rispondenti evidenziano che i principali sono: valutare la preparazione sugli argomenti studiati, accedere a chiarimenti su argomenti ritenuti difficili o dubbi, accedere a materiali di approfondimento. Meno avvertiti sono i vantaggi, corrispondenti a funzionalità più avanzate dell’Intelligenza Artificiale, come: la possibilità di avere un compagno di studio e la possibilità di avere uno strumento che si adatti allo stato emotivo della persona. Coerentemente con quanto già dichiarato, tra le specifiche funzioni ritenute importanti, le rispondenti indicano principalmente l’elaborazione e la classificazione di elevate quantità di dati e informazioni e la raccomandazione di contenuti o consigli di approfondimento; la funzione meno indicata è la possibile presa di decisioni da parte dell’Intelligenza Artificiale indipendentemente dall’utente. Tra le paure e resistenze nell’utilizzo dell’Intelligenza Artificiale nel percorso formativo, le rispondenti indicano principalmente la perdita del contatto umano con docenti e altri studenti e l’incapacità da parte di tali sistemi di abilitare una *vera* inclusione; tra le paure meno indicate, ci sono la mancanza di trasparenza generale dei processi e la preoccupazione circa l’impiego dei dati personali. Tra le caratteristiche che potrebbero rappresentare punti di forza che spingerebbero ad utilizzare sistemi

di Intelligenza Artificiale in ambito educativo, viene indicata principalmente la possibilità data all'utente di decidere come usare tali sistemi e quando, nonché di accedere a spiegazioni sulla modalità di utilizzo e funzionamento della tecnologia. I cambiamenti ritenuti necessari affinché si implementino sistemi di Intelligenza Artificiale in ambito educativo riguardano la preparazione di tutti gli attori coinvolti nel processo formativo: non solo gli studenti, ma anche i docenti e l'università come istituzione. Infine, in generale, le studentesse evidenziano una media disponibilità ad applicare sistemi di Intelligenza Artificiale nei propri processi di apprendimento.

3. Le potenzialità dell'applicazione di un approccio human-centered

I risultati preliminari di questa survey indicano che, sebbene i sistemi di Intelligenza Artificiale vengano impiegati da parte di studenti e studentesse, che in generale si pongono in maniera mediamente positiva nei confronti di tale argomento, le funzionalità impiegate sono ancora limitate, mentre dubbi e paure sono numerose. Considerato che l'impiego dell'Intelligenza Artificiale nel settore educativo sembra ormai inevitabile, è necessario guidare tale processo. L'idea che tale tecnologia renda tutto più semplice e immediato dovrebbe, infatti, cedere il passo ad un approccio più consapevole e critico che porti ad un'innovazione più complessa del settore educativo universitario. Per tale ragione, è anzitutto necessario il coinvolgimento di tutti gli stakeholder coinvolti, dal livello micro (gli studenti) a quello meso (docenti), fino a quello macro (enti di formazione e istituzioni responsabili delle politiche pubbliche). L'applicazione dell'approccio dello *human-centered design* (IxDF, 2021; Lupton, 2018) può essere fondamentale in

quest'ottica. Tale approccio, il cui obiettivo è quello di progettare soluzioni (digitali e non) per sistemi complessi ponendo al centro i bisogni, le esperienze e le aspettative delle persone, potrebbe, infatti, essere usato nel *co-design* di sistemi che mirino a:

- creare interfacce e sistemi comprensibili e accessibili, ad esempio fornendo tutorial integrati e guide interattive per supportare l'apprendimento delle funzionalità;
- progettare sistemi flessibili che permettano un alto grado di personalizzazione (es.: regolare le impostazioni e ricevere suggerimenti), in modo da aumentare la percezione di controllo e rendere l'uso della tecnologia più conforme alle preferenze individuali;
- creare sistemi che integrino le interazioni umane (ad esempio con i docenti) piuttosto che allontanarle, potenziando laddove possibile la comunicazione, piuttosto che sostituirla;
- creare sistemi accessibili per persone con diverse abilità e background socio-culturali;
- creare strumenti che forniscano feedback personalizzati e valutazioni adattive.

Si tratta di esempi elaborati sulla base dei risultati preliminari dello studio. Una volta conclusa l'analisi si organizzeranno sessioni di *co-design* utili alla progettazione di un sistema di Intelligenza Artificiale basato sui bisogni dei diversi attori coinvolti.

Bibliografia

Interaction Design Foundation - IxDF. (2021), *What is Human-Centered Design (HCD)?*, Interaction Design Foundation - IxDF. <https://www.interaction-design.org/literature/topics/human-centered-design>.

Lupton D. (2018), Towards Design Sociology. *Sociology Compass*, 12(1), 1-11.

Opromolla A. (2024), *Come l'IA cambia l'educazione: strategie, ricerca, nodi etici*. Agenda Digitale, <https://www.agenda-digitale.eu/scuola-digitale/come-lia-cambia-leducazione-strategie-ricerca-nodi-etici/>.

RUBRICA *EMPOWERMENT*

1. APPROCCIO UMANOCENTRICO ALL'INTELLIGENZA ARTIFICIALE: SFIDE ETICHE, SOCIALI ED ECONOMICHE

di Riccardo Mancini, Sara Pellegrini, Riccardo Sebastiani, Debora Glori*

Abstract: L'articolo esplora un approccio umanocentrico all'intelligenza artificiale (IA), indagandone implicazioni etiche, pedagogiche e sociali attraverso le percezioni degli studenti del TFA. L'obiettivo è definire linee guida per un utilizzo sostenibile e responsabile dell'IA che integri valori fondamentali, promuova il benessere educativo e sociale e risponda criticamente alle sfide poste dall'automazione.

Parole chiave: Umanocentrico, Bias, Governance, Sostenibilità, Etica.

Abstract: The article explores a human-centric approach to artificial intelligence (AI), examining its ethical, pedagogical, and social implications through the perceptions of TFA students. The objective is to define guidelines for a sustainable and responsible use of AI that integrates fundamental values, promotes educational and social well-being, and critically addresses the challenges posed by automation.

Keywords: Human-centric, Bias, Governance, Sustainability, Ethics.

* Riccardo Mancini, Dipartimento di Scienze Umane, Università degli Studi "Link", r.mancini@unilink.it. Sara Pellegrini, Dipartimento di Scienze Umane, Università degli Studi "Link", s.pellegrini@unilink.it. Riccardo Sebastiani, Dipartimento di Scienze Umane, Università degli Studi "Link", r.sebastiani@unilink.it. Debora Glori, Università degli Studi "Link", MIM, debora.glori97@gmail.com.

Accettato Dicembre 2024 - Pubblicato Aprile 2025.

Introduzione

L'Intelligenza Artificiale (IA) ha cambiato il panorama tecnologico e sociale in maniera irreversibile. Si parla di “quarta rivoluzione industriale” e delle sue implicazioni etiche, sociali ed economiche, soprattutto nell’ottica di un approccio umanocentrico, che mira a integrare valori universali come il rispetto, la giustizia e la sostenibilità. Il nostro studio esplora come l’IA, se guidata da principi etici solidi, possa contribuire al benessere collettivo.

L’approccio umanocentrico pretende di mettere l’essere umano al centro delle tecnologie emergenti, riconoscendone la dignità e i diritti, implicando lo sviluppo di tecnologie efficaci e soprattutto di quelle giuste e trasparenti.

Dalle sue origini negli anni ‘50, l’IA si è evoluta, guadagnando fiducia e scetticismo. Le persone hanno percezioni contrastanti: alcuni vedono l’IA come il “cavaliere” della nuova era tecnologica, mentre altri temono la perdita di controllo e autonomia. L’Unione Europea ha adottato un ruolo normativo e regolatorio di primaria importanza nel contesto dello sviluppo e dell’applicazione dell’intelligenza artificiale. Piuttosto che sviluppare direttamente sistemi di IA, l’UE si concentra sulla definizione di un quadro giuridico e deontologico volto a garantire un impiego responsabile e orientato al bene comune di tali tecnologie. Attraverso strumenti legislativi come l’AI Act, essa promuove standard di trasparenza, equità e sicurezza, con l’obiettivo di armonizzare l’adozione dell’IA nei paesi membri in un’ottica di tutela dei diritti fondamentali e del progresso socio-economico. Le percezioni pubbliche influenzano fortemente la regolamentazione e le politiche dell’IA e questo riflette sulle credenze delle persone. Secondo uno studio recente, i cittadini europei mostrano grande preoccupazione per la privacy e la trasparenza nelle decisioni algoritmiche. Le implicazioni etiche

dell'IA sono complesse e spesso problematiche. I bias algoritmici, ad esempio, derivano da dati “viziati” che, se non bilanciati, perpetuano disuguaglianze. Alcuni sistemi di riconoscimento facciale hanno mostrato tassi di errore elevati per gruppi di minoranze etniche e questo solleva domande urgenti su equità e giustizia. Inoltre, la raccolta di grandi quantità di dati personali fa emergere preoccupazioni per la privacy e l'autonomia dell'individuo. Un famoso esempio di bias algoritmico è il sistema COMPAS, utilizzato nel sistema giudiziario statunitense per valutare il rischio di recidiva. Uno studio condotto da Angwin *et al.* (2016) per ProPublica ha dimostrato che COMPAS tendeva a sovrastimare il rischio per le minoranze afroamericane e a sottostimarne per i bianchi, contribuendo a decisioni discriminatorie nel sistema penale. Questo caso ha sollevato questioni fondamentali sull'opacità e sulla responsabilità degli algoritmi decisionali, evidenziando la necessità di regolamentazioni più stringenti in materia di bias e trasparenza.

L'intelligenza artificiale è oggi un terreno di competizione geopolitica tra Unione Europea, Stati Uniti e Cina. Mentre gli Stati Uniti guidano l'innovazione tecnologica grazie a colossi come Google, Microsoft e OpenAI, la Cina ha investito massicciamente in IA attraverso il supporto statale a imprese come Baidu, Alibaba e Tencent. L'Unione Europea, invece, si distingue per il suo approccio regolamentare, come dimostrato dall'AI Act, volto a garantire un uso etico e sicuro dell'IA nei paesi membri. Questa diversità di approcci pone questioni fondamentali sulla sovranità digitale e sulla governance globale dell'intelligenza artificiale.

Il problema dell'automazione nel mercato del lavoro non può essere trattato in modo generico. Secondo il rapporto McKinsey (2022), settori come la manifattura, la logistica e i servizi di customer service saranno i più colpiti dall'automazione, mentre professioni ad alta specializzazione cognitiva e creativa, come

quelle nel settore sanitario e nell'ingegneria, avranno una maggiore resilienza. Studi dell'OCSE (2021) indicano che entro il 2030 il 14% dei posti di lavoro potrebbe essere automatizzato, mentre un ulteriore 32% subirà trasformazioni significative. Questo scenario sottolinea la necessità di programmi di riqualificazione per la forza lavoro e politiche di welfare per mitigare le disuguaglianze economiche.

Una governance dell'IA è essenziale per bilanciare innovazione e responsabilità. L'Unione Europea ha proposto un quadro normativo basato su trasparenza e accountability, ma la sfida è mantenere le norme sufficientemente flessibili per seguire i rapidi avanzamenti tecnologici. Diverse organizzazioni internazionali stanno esplorando il concetto di una "governance adattativa", che prevede la collaborazione tra settori pubblico e privato e tra governi e cittadini.

«La governance adattativa risponde all'esigenza di una regolamentazione "dinamica", che si aggiorni in tempo reale alle innovazioni e ai rischi emergenti, un concetto promosso anche dall'OECD nel 2023» (OECD, 2023).

L'IA potrebbe giocare un ruolo chiave nel promuovere la sostenibilità ecologica, migliorando ad esempio l'efficienza energetica e la gestione delle risorse. Tuttavia, i modelli di IA, specialmente quelli più complessi, richiedono risorse significative. È quindi cruciale adottare una prospettiva di "ecologia integrale", considerare l'interazione tra IA, ambiente e comunità. La ricerca su algoritmi green e a basso consumo energetico è in crescita e rappresenta un'opportunità per uno sviluppo tecnologico rispettoso dell'ambiente.

«La ricerca ha dimostrato che l'addestramento di modelli di deep learning avanzati può avere un'impronta di carbonio simile a quella di un'auto nel suo ciclo di vita» (TECH4FUTURE, 2022).

Il potenziale dell'IA nell'educazione è immenso. I metodi di apprendimento personalizzati, resi possibili dai sistemi di IA, migliorano la didattica e favoriscono un apprendimento mirato. Tuttavia, è necessario evitare una dipendenza dalla tecnologia che possa portare a disuguaglianze educative. Un approccio consapevole è fondamentale per assicurare che l'IA supporti gli insegnanti e non li sostituisca. Secondo l'UNESCO, la tecnologia dovrebbe essere integrata nei curricula educativi con una formazione specifica per insegnanti e studenti, per massimizzare i benefici e ridurre i rischi di dipendenza digitale (UNESCO, 2022). Con l'automazione in rapida espansione, molte occupazioni tradizionali rischiano di essere sostituite, portando a una riorganizzazione dei settori lavorativi. Le politiche di riqualificazione sono essenziali per preparare la forza lavoro a nuove competenze. Secondo l'OCSE, entro il 2030, il 15% dei lavoratori potrebbe dover adattarsi a nuovi ruoli a causa dell'automazione. Uno studio di McKinsey (2022) ha rilevato che la collaborazione tra datori di lavoro, istituzioni educative e governi è fondamentale per preparare la forza lavoro ai cambiamenti futuri.

1. Un approccio umanocentrico

L'approccio umanocentrico all'intelligenza artificiale è oggi ritenuto fondamentale per garantire uno sviluppo tecnologico che rispetti i principi etici e i valori umani. Questo tipo di approccio richiede che la progettazione e l'implementazione dell'IA siano guidate da una visione che pone al centro l'essere umano, promuovendo trasparenza e giustizia in ogni fase del ciclo di vita dell'algoritmo. La trasparenza, infatti, è essenziale per rendere comprensibili i processi decisionali automatizzati e per mitigare il rischio di esiti ingiusti o discriminatori (Floridi & Cowls, 2020).

Implementare sistemi IA che siano non solo tecnicamente avanzati ma anche spiegabili e tracciabili consente una maggiore fiducia da parte degli utenti e delle comunità, evitando la cosiddetta “scatola nera” degli algoritmi. La partecipazione inclusiva di tutti gli stakeholder rappresenta un altro pilastro cruciale. Coinvolgere un’ampia gamma di attori – tra cui accademici, sviluppatori, responsabili politici e rappresentanti della società civile – garantisce che le decisioni sulla progettazione e l’utilizzo dell’IA riflettano una molteplicità di prospettive e siano rappresentative delle esigenze della società (Johnson & Verdicchio, 2022). Questa forma di partecipazione favorisce processi decisionali più equi e inclusivi, evitando che le tecnologie emergenti diventino strumenti di esclusione o discriminazione sociale.

Un’attenzione particolare deve essere rivolta anche all’educazione digitale. L’alfabetizzazione critica sull’IA non si limita alla formazione tecnica, ma si estende alla capacità di comprendere e riflettere sulle implicazioni etiche e sociali della tecnologia (Smith, 2023). Educare le nuove generazioni e riqualificare la forza lavoro esistente attraverso programmi mirati può stimolare una cultura della consapevolezza e della responsabilità collettiva. Investire in percorsi educativi che sviluppino non solo competenze tecniche, ma anche capacità di analisi e pensiero critico, è fondamentale per rendere l’IA un alleato e non una minaccia. Infine, la regolamentazione deve essere flessibile e adattabile ai rapidi cambiamenti del panorama tecnologico, senza tuttavia sacrificare i principi fondamentali di rispetto e dignità umana. Una regolamentazione dinamica e adattabile può garantire un equilibrio tra l’innovazione e la protezione dei diritti umani, promuovendo un uso responsabile e etico dell’IA. La governance inclusiva e collaborativa, che integra approcci transdisciplinari e coordina le risorse globali, può trasformare l’IA in uno strumento che sia capace di risolvere

problemi complessi ma lo fa mantenendo una prospettiva etica e umana (Floridi & Cows, 2020; Johnson & Verdicchio, 2022).

La complessità spinge ad intraprendere un approccio che bilancia progresso tecnologico e valori umani per costruire un ecosistema in cui l'IA serva la collettività: l'adozione di norme che promuovono la giustizia, l'equità e la sostenibilità è una necessità per realizzare uno strumento al servizio del benessere comune.

Un approccio umanocentrico all'intelligenza artificiale richiede una sinergia tra discipline diverse, creando un dialogo tra scienze tecniche e saperi umanistici. Questo tipo di collaborazione multidisciplinare è essenziale per garantire che i principi etici e i valori umani siano integrati a tutti i livelli dello sviluppo e dell'applicazione dell'IA. Filosofia, diritto, psicologia e scienze comportamentali devono lavorare fianco a fianco con ingegneri e data scientist per costruire un'IA che non sia solo tecnologicamente avanzata, ma anche profondamente rispettosa delle diversità e delle esigenze sociali (Floridi, 2020). L'interazione tra queste discipline permette di affrontare le sfide etiche e sociali dell'IA con un approccio che va oltre la pura funzionalità, includendo considerazioni su equità, inclusione e giustizia. La filosofia etica, ad esempio, gioca un ruolo fondamentale nel mettere in discussione cosa significhi davvero creare un sistema giusto e orientato al bene comune in un contesto dominato dall'automazione. Le scienze sociali, invece, forniscono una lente attraverso cui esaminare l'impatto dell'IA sulle dinamiche sociali e sui rapporti di potere, aiutando a individuare le potenziali disuguaglianze che potrebbero essere amplificate dalla tecnologia. Le scienze comportamentali offrono un ulteriore contributo, esplorando come le persone interagiscono con i sistemi intelligenti e come i bias cognitivi possano influenzare le decisioni umane durante l'uso di queste tecnologie. La regolamentazione gioca un ruolo chiave in questo quadro, non solo come strumento di

controllo ma come mezzo per costruire una fiducia condivisa. Le leggi e le normative devono essere proattive, capaci di anticipare e affrontare le sfide etiche con un approccio che non si limiti a reagire ai problemi ma li prevenga (Smith, 2023). Solo attraverso una regolamentazione pensata per proteggere i diritti umani e promuovere la trasparenza è possibile garantire che l'IA rimanga uno strumento di progresso collettivo. Integrare le competenze delle diverse discipline crea una base solida per sviluppare tecnologie che non solo rispondono alle esigenze tecniche ma abbracciano anche le complessità etiche e sociali del nostro tempo. Questa integrazione trasforma l'IA in un alleato dell'umanità, capace di costruire un futuro più giusto, inclusivo e sostenibile.

2. Metodologia di ricerca

Questo studio si avvale di un approccio metodologico misto, combinando strumenti di ricerca qualitativi e quantitativi per analizzare l'impatto dell'intelligenza artificiale (IA) in ambiti sociali, etici ed economici. L'obiettivo è esaminare come l'IA venga percepita, regolamentata e applicata in diversi settori, valutandone le implicazioni a lungo termine, con particolare attenzione alle dinamiche di accettazione, regolamentazione e impatto sulle professioni. La ricerca è stata strutturata in più fasi per garantire una raccolta dati rigorosa e approfondita. In primo luogo, è stata effettuata una revisione della letteratura scientifica esistente, finalizzata a individuare i principali quadri teorici e le prospettive critiche in merito all'IA. Successivamente, la raccolta dati è stata condotta attraverso questionari strutturati e interviste semi-strutturate. Il questionario, elaborato con la piattaforma SurveyMonkey, è stato distribuito a un campione eterogeneo di partecipanti appartenenti a diversi ambiti professionali e

accademici, al fine di raccogliere dati su conoscenze, preoccupazioni, e aspettative rispetto all'IA. Le domande hanno indagato aspetti quali il livello di familiarità con l'IA, il grado di fiducia nelle sue applicazioni e le implicazioni etiche percepite. È stato selezionato un campione rappresentativo di 227 insegnanti delle scuole pubbliche italiane, utilizzando un campionamento stratificato per area geografica e tipo di scuola. La rappresentatività è stata garantita attraverso la proporzionalità tra i sottogruppi e la popolazione complessiva di riferimento con un margine di errore del 5%. Per garantire un'analisi più approfondita, le interviste semi-strutturate sono state condotte con un sottoinsieme selezionato di partecipanti, scelti in base alla loro esperienza professionale e accademica nel settore dell'IA.

Le interviste hanno permesso di esplorare in dettaglio le percezioni, le esperienze dirette e le aspettative degli intervistati, integrando così i dati quantitativi con prospettive qualitative più articolate. L'analisi quantitativa delle risposte ai questionari è stata effettuata mediante tecniche statistiche descrittive e inferenziali, al fine di individuare correlazioni significative tra le variabili considerate. I dati qualitativi, derivanti dalle interviste, sono stati analizzati attraverso un approccio tematico, che ha permesso di identificare ricorrenze e tendenze emergenti, nonché eventuali divergenze nelle percezioni tra diverse categorie di partecipanti. Un elemento chiave della ricerca è stato il confronto tra gruppi di intervistati, al fine di comprendere come variabili quali età, professione, settore di appartenenza e livello di istruzione influiscano sulla percezione dell'IA. Tale confronto ha consentito di delineare profili distinti di atteggiamenti nei confronti dell'IA, evidenziando differenze tra categorie di utenti più esperti e utenti con minore familiarità con la tecnologia. Per garantire l'affidabilità e la validità dei dati raccolti, è stato implementato un processo di validazione interna. Un campione pilota ha testato la coerenza e la

comprensibilità del questionario prima della somministrazione su larga scala. Inoltre, le interviste sono state condotte seguendo criteri standardizzati, con domande aperte finalizzate a garantire una comparabilità tra le risposte. Infine, l'analisi dei dati ha consentito di individuare le principali sfide e opportunità legate all'IA, mettendo in evidenza il bisogno di regolamentazioni adeguate, percorsi di alfabetizzazione digitale e strategie di governance efficaci per favorire un uso etico e responsabile delle tecnologie emergenti. L'integrazione tra dati quantitativi e qualitativi ha reso possibile una visione più completa e articolata del fenomeno, permettendo di avanzare raccomandazioni informate per la gestione dell'IA in ambito sociale ed economico.

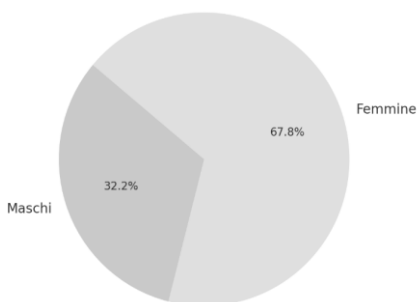
3. Risultati

L'analisi dei dati raccolti ha rivelato tendenze significative in merito alle percezioni e alle preoccupazioni riguardanti l'adozione dell'intelligenza artificiale in diversi settori disciplinari. I partecipanti, che includevano una vasta gamma di età e background accademici e professionali, hanno espresso opinioni variegata, ma emergono alcune osservazioni comuni. Hanno risposto all'indagine 227 partecipanti. Essi sono così distribuiti: n= 227, di cui 73 maschi e 154 femmine, come nel grafico 1. L'intero campione si colloca all'interno di una fascia di età che va dai 20 anni ai 59 anni, come nel grafico 2. La maggior parte degli intervistati ha indicato la medicina e l'ingegneria come i campi più influenzati dall'IA, un risultato che sottolinea la percezione diffusa dell'impatto tecnologico su settori orientati sia alla cura della salute sia allo sviluppo tecnologico avanzato. In termini di interessi specifici, l'apprendimento automatico e l'impatto economico dell'IA sono stati temi molto discussi, evidenziando come gli intervistati

riconoscano l'importanza delle innovazioni tecniche insieme alle loro implicazioni finanziarie. Tra le preoccupazioni più frequenti spiccano la potenziale sostituzione del lavoro umano e il rischio di controllo eccessivo, aspetti che mettono in luce il timore di una progressiva perdita di controllo e di una dipendenza crescente dai sistemi automatizzati. Una parte rilevante dell'analisi ha riguardato la risposta etica all'uso dell'IA, con molti partecipanti che hanno dichiarato che l'uso della tecnologia nella prevenzione delle malattie è accettabile solo a determinate condizioni. Questo indica una riflessione etica complessa e diversificata, dove la fiducia nella tecnologia è bilanciata da considerazioni morali. Infine, la probabilità con cui i partecipanti raccomanderebbero l'uso dell'IA a conoscenti o colleghi ha mostrato una gamma di risposte che vanno da livelli di entusiasmo moderato a posizioni più caute, riflettendo la necessità di ulteriore trasparenza e regole etiche per rafforzare la fiducia pubblica.

Grafico 1: La demografia

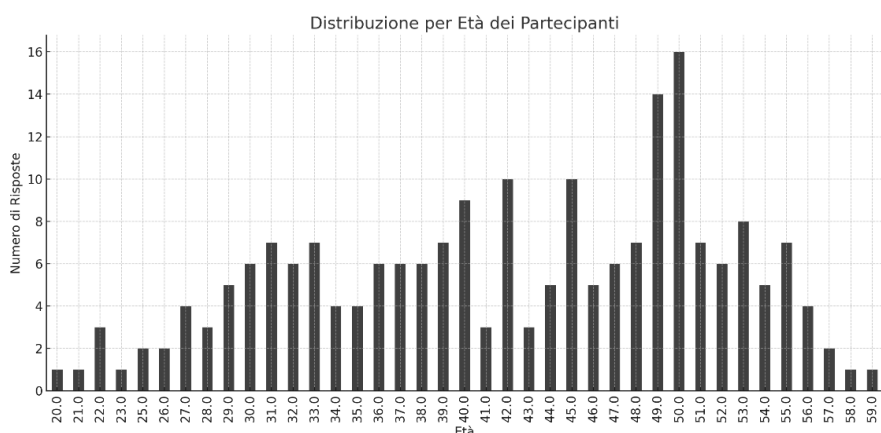
Distribuzione dei partecipanti per genere (n=227)



Il grafico 2 che rappresenta la distribuzione per età dei partecipanti al sondaggio evidenziando una partecipazione equamente distribuita tra diverse fasce d'età, con ciascuna categoria rappresentata da un numero uniforme di risposte. La

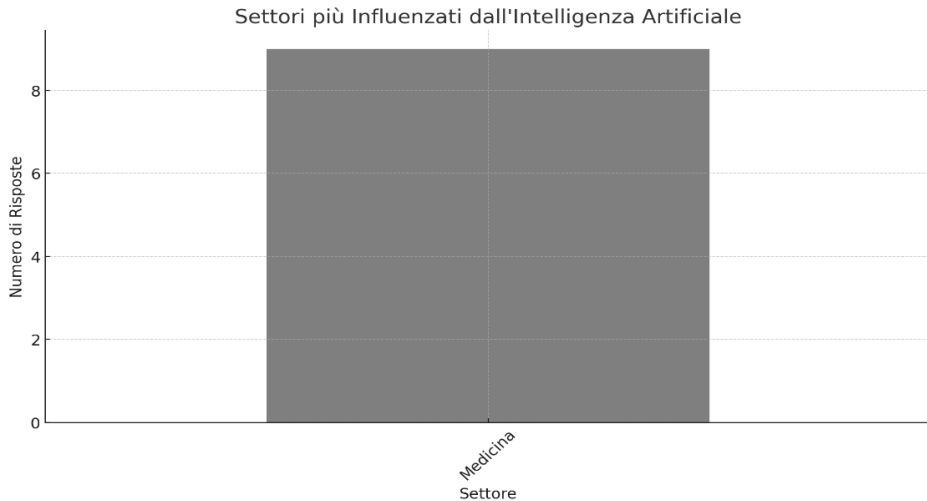
varietà demografica assicura una visione bilanciata delle opinioni raccolte, fornendo un quadro eterogeneo delle percezioni sull'impatto dell'intelligenza artificiale in vari settori. La partecipazione intergenerazionale è importante per comprendere come le diverse età percepiscano e si relazionino alla tecnologia. I più giovani potrebbero avere un approccio più flessibile e ottimista, essendo cresciuti in un contesto tecnologico avanzato, mentre le persone di età superiore potrebbero mostrare una maggiore cautela o scetticismo, legata a una conoscenza più tradizionale e a un'esperienza professionale consolidata. Questa varietà di prospettive contribuisce a delineare un panorama più completo delle aspettative, preoccupazioni e potenziali barriere legate all'adozione dell'IA. L'assenza di picchi significativi in una fascia d'età specifica indica che il tema trattato è di rilevanza trasversale e interessa diverse generazioni. Questa distribuzione omogenea dei partecipanti suggerisce che l'IA è un argomento che coinvolge un ampio spettro della popolazione, sia in termini di interesse accademico che professionale, confermando la sua crescente importanza e pervasività nei contesti moderni.

Grafico 2: I partecipanti



Il grafico 3 mostra chiaramente che la “Medicina” è percepita come il settore maggiormente influenzato dall’intelligenza artificiale. Il numero di risposte, rappresentato sul grafico, indica un forte consenso tra i partecipanti riguardo al ruolo centrale che l’IA ha assunto nella trasformazione dei processi e delle pratiche mediche. Questa tendenza suggerisce che la comunità riconosce l’IA non solo come un supporto alle decisioni cliniche, ma anche come uno strumento capace di migliorare la diagnostica, personalizzare i trattamenti e ottimizzare le operazioni sanitarie. L’attenzione particolare verso la medicina può essere attribuita ai recenti progressi tecnologici che hanno integrato l’IA con tecnologie avanzate come l’analisi dei big data e l’apprendimento automatico. Questi sviluppi permettono di gestire una mole crescente di informazioni mediche con maggiore precisione e rapidità, migliorando così l’efficienza e l’efficacia dei trattamenti. Il dato riflette anche una crescente consapevolezza del potenziale dell’IA per affrontare le sfide del settore sanitario, come la necessità di diagnosi precoci e l’ottimizzazione delle risorse cliniche. Il fatto che nessun altro settore sia rappresentato nel grafico con numeri significativi potrebbe indicare una percezione condivisa che l’IA abbia un impatto particolarmente incisivo e immediato sulla medicina rispetto ad altre discipline. Questa concentrazione di risposte sulla medicina sottolinea l’importanza di politiche mirate e una regolamentazione attenta che garantisca un utilizzo etico dell’IA, rispettando la privacy dei pazienti e mitigando i rischi legati all’automazione. In conclusione, il grafico evidenzia un punto di vista unanime su come la tecnologia stia ridefinendo la pratica medica e contribuisca a porre le basi per un futuro sanitario più efficiente e avanzato.

Grafico 3: I Settori principali



4. Discussione

L'analisi approfondita dei dati emersi dall'indagine offre un quadro chiaro e sfaccettato delle percezioni sull'impatto dell'intelligenza artificiale (IA) nei vari settori disciplinari, oltre a riflettere le opinioni e le preoccupazioni dei partecipanti rispetto alla sua applicazione pratica ed etica. La medicina è emersa come il settore principale percepito come più influenzato dall'IA, una constatazione che riflette i recenti progressi nell'integrazione della tecnologia con la pratica clinica. L'IA, infatti, è sempre più utilizzata per migliorare la diagnosi precoce, personalizzare i piani terapeutici e ottimizzare le operazioni sanitarie, contribuendo ad un'evoluzione della medicina verso un approccio più preciso e basato sui dati. L'ingegneria è un altro settore rilevante, visto come profondamente trasformato dall'IA, sia nella progettazione che nell'automazione dei processi. Le risposte ricevute indicano che i

partecipanti riconoscono l'IA come una forza trainante per l'innovazione tecnologica, capace di aumentare l'efficienza e ridurre i margini di errore. Tuttavia, il quadro non è privo di ombre: una parte consistente dei partecipanti ha evidenziato le sfide legate alla transizione tecnologica, sottolineando le implicazioni etiche e sociali che derivano dall'automazione avanzata e dalla crescente dipendenza dai sistemi intelligenti.

Tra i temi più discussi emergono anche le preoccupazioni legate alla sostituzione del lavoro umano. Secondo il rapporto McKinsey (2022), settori come la manifattura, la logistica e i servizi di customer service saranno i più colpiti dall'automazione, mentre professioni ad alta specializzazione cognitiva e creativa, come quelle nel settore sanitario e nell'ingegneria, avranno una maggiore resilienza. Studi dell'OCSE (2021) indicano che entro il 2030 il 14% dei posti di lavoro potrebbe essere automatizzato, mentre un ulteriore 32% subirà trasformazioni significative. Questo scenario sottolinea la necessità di programmi di riqualificazione per la forza lavoro e politiche di welfare per mitigare le disuguaglianze economiche. Questo aspetto è particolarmente rilevante nei contesti in cui l'IA potrebbe sostituire lavori tradizionalmente eseguiti dall'uomo, con implicazioni dirette per la forza lavoro e per la stabilità economica di diverse comunità. Le risposte indicano la necessità di politiche che bilancino l'adozione dell'IA con misure di protezione sociale e piani di riqualificazione professionale per sostenere la forza lavoro nella transizione. Un'altra area di discussione importante riguarda l'etica dell'uso dell'IA, soprattutto in settori sensibili come la prevenzione delle malattie. Molti partecipanti hanno espresso opinioni articolate sull'accettabilità etica dell'IA in ambito sanitario, con alcuni che ritengono accettabile l'uso della tecnologia solo in presenza di una forte trasparenza e regolamentazione. Questi dati evidenziano come, nonostante il potenziale dell'IA di migliorare la salute pubblica e

l'efficienza dei sistemi sanitari, persistano dubbi legati alla protezione della privacy, alla gestione dei dati sensibili e al rispetto del consenso informato. L'analisi dei dati ha anche messo in evidenza come la propensione a raccomandare l'uso dell'IA vari considerevolmente tra i partecipanti. Questa variabilità suggerisce che, mentre alcuni vedono nell'IA un'opportunità per migliorare la qualità della vita e semplificare processi complessi, altri rimangono scettici, preoccupati per gli effetti collaterali che potrebbero derivare da un'adozione indiscriminata della tecnologia. Le differenze di opinione possono essere attribuite a diversi livelli di conoscenza e familiarità con la tecnologia, nonché alle esperienze personali e professionali degli intervistati. Questa variabilità suggerisce che, mentre alcuni vedono nell'IA un'opportunità per migliorare la qualità della vita e semplificare processi complessi, altri rimangono scettici, preoccupati per gli effetti collaterali che potrebbero derivare da un'adozione indiscriminata della tecnologia. Infine, il tema dell'impatto culturale dell'IA è emerso come un punto di riflessione significativo. Alcuni partecipanti hanno sottolineato il rischio di un appiattimento culturale e di una perdita di capacità critiche e creative, a fronte di un uso eccessivo dell'IA per automatizzare attività cognitive. Questa visione porta alla luce una preoccupazione più ampia: come mantenere l'equilibrio tra l'adozione dell'IA e la preservazione delle competenze umane essenziali, come il pensiero critico e la creatività, che costituiscono le fondamenta della cultura e dell'innovazione. I dati raccolti offrono una visione dettagliata e complessa dell'impatto dell'IA, evidenziando il bisogno di una strategia di adozione equilibrata e consapevole.

L'intelligenza artificiale rappresenta una forza trasformativa con un potenziale straordinario, ma la sua applicazione richiede una riflessione continua sulle implicazioni etiche, economiche e sociali per garantire che sia utilizzata come

strumento di progresso inclusivo. Dall'analisi dei dati emersi, è evidente che l'intelligenza artificiale (IA) rappresenta un potente strumento con numerosi punti di forza che ne favoriscono l'adozione in vari settori. Tra questi, spiccano l'incremento dell'efficienza operativa e la capacità di elaborare grandi quantità di dati con rapidità e precisione. In particolare, settori come la medicina e l'ingegneria beneficiano di strumenti avanzati di apprendimento automatico e reti neurali profonde, che migliorano la diagnostica, ottimizzano i processi e riducono gli errori umani. La capacità dell'IA di personalizzare i trattamenti e prevedere scenari complessi contribuisce a una maggiore precisione e a un miglioramento generale della qualità delle prestazioni. Un altro punto di forza significativo è la sua potenzialità di alleggerire i carichi di lavoro e migliorare l'efficienza aziendale. I partecipanti all'indagine hanno riconosciuto che, se ben implementata, l'IA può ridurre il tempo dedicato a compiti ripetitivi e amministrativi, liberando risorse umane per attività più strategiche e creative. Questa prospettiva rende l'IA un alleato per aumentare la produttività e stimolare l'innovazione. Nonostante i vantaggi, l'IA presenta anche punti deboli significativi. Tra le preoccupazioni principali sollevate dai partecipanti, emerge il rischio di sostituzione del lavoro umano e la conseguente disoccupazione. La paura che l'automazione possa erodere la stabilità economica di vari settori lavorativi è un punto di debolezza che richiede attenzione e misure di mitigazione, come la riqualificazione e l'aggiornamento delle competenze professionali. Inoltre, sono state sollevate questioni etiche legate all'uso dell'IA, specialmente in ambiti delicati come la medicina, dove la privacy dei dati e il consenso informato sono temi cruciali. Un'altra criticità evidenziata è la possibilità che l'uso eccessivo dell'IA possa portare a una passività intellettuale, limitando la capacità delle persone di pensare in modo critico e creativo. Questo appiattimento delle

competenze umane rischia di impoverire la cultura e ridurre la diversità delle prospettive, soprattutto se l'IA viene adottata senza un'adeguata supervisione e regolamentazione.

Guardando al futuro, le prospettive per l'IA sono promettenti, ma richiedono un approccio equilibrato e strategico. L'implementazione dell'IA deve essere accompagnata da regolamentazioni solide e da una governance che assicuri l'uso etico e responsabile della tecnologia. L'educazione e la formazione continua saranno essenziali per preparare le future generazioni a convivere con l'IA, sfruttando i suoi vantaggi senza perdere di vista l'importanza delle capacità umane come il pensiero critico e la creatività. Le politiche future dovrebbero focalizzarsi sull'integrazione dell'IA in modo che sia complementare alle attività umane, promuovendo l'innovazione senza sacrificare la dimensione umana del lavoro e della società. Investimenti in ricerca interdisciplinare, che uniscano tecnologia, etica e scienze sociali, saranno fondamentali per costruire un ecosistema in cui l'IA possa prosperare come strumento al servizio dell'umanità, mitigando i rischi e massimizzando i benefici. In definitiva, per realizzare appieno il potenziale dell'IA, sarà essenziale un dialogo costante tra sviluppatori, regolatori e la società civile per assicurare che questa tecnologia rimanga un potente alleato nel progresso umano, capace di rispondere alle esigenze di un mondo sempre più interconnesso e complesso.

Conclusioni

Le conclusioni di questo saggio riflettono l'importanza di un approccio umanocentrico all'intelligenza artificiale per affrontare le sfide etiche, sociali ed economiche legate alla sua adozione. L'analisi condotta dimostra che, sebbene l'IA offra

innegabili vantaggi in termini di innovazione e progresso tecnologico, la sua implementazione non può prescindere da una solida cornice di principi etici e valori umani. La medicina emerge come esempio chiave di come l'IA possa migliorare la qualità della vita e l'efficienza dei servizi, ma al contempo solleva interrogativi su privacy, equità e giustizia. Per massimizzare i benefici dell'IA e minimizzare i rischi, è fondamentale sviluppare un modello di governance dinamico che promuova trasparenza, inclusione e responsabilità. La collaborazione tra le discipline e il dialogo tra settore pubblico e privato può favorire un'integrazione armoniosa dell'IA nella società, mantenendo al centro l'interesse umano. Inoltre, l'educazione e la formazione critica devono essere prioritarie, così da preparare le future generazioni a utilizzare e comprendere l'IA in modo consapevole e responsabile. Infine, la ricerca sottolinea che il potenziale dell'IA per contribuire a un futuro sostenibile e inclusivo dipende dall'impegno collettivo per regolamentare e guidare il suo sviluppo in una direzione etica e orientata al bene comune. Solo attraverso un approccio equilibrato, che bilanci l'innovazione con il rispetto dei diritti e della dignità umana, l'IA potrà realizzare il suo potenziale di motore di progresso per tutta l'umanità.

Bibliografia

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks*. ProPublica.

Bostrom, N. (2018). *Superintelligenza: Tendenze, pericoli, strategie* (S. Frediani, Trad.). Bollati Boringhieri. (Opera originale pubblicata nel 2014).

Brynjolfsson, E., & McAfee, A. (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W.W. Norton & Company.

Coeckelbergh, M. (2020). *AI Ethics*. MIT Press.

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press.

European Commission. (2023). *White Paper on Artificial Intelligence: A European Approach to Excellence and Trust*.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. C., & Srikumar, M. (2020). *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI*. Berkman Klein Center Research Publication, (2020-1).

Floridi, L., (2020). *The Ethics of Artificial Intelligence: Principles and Practices*. Oxford University Press.

Floridi, L., (2022). *Etica dell'intelligenza artificiale*. Raffaello Cortina.

Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines*, 30(1), 99-120.

Jobin, A., Ienca, M., & Vayena, E. (2020). The Global Landscape of AI Ethics Guidelines. *Nature Machine Intelligence*, 2(6), 389-399.

Johnson, M., & Verdicchio, M. (2022). *AI Literacy and Ethical Challenges in the 21st Century*. Cambridge Scholars Publishing.

McKinsey Global Institute. (2022). *The Future of Work in the Age of AI*.

Mittelstadt, B. D. (2021). Principles Alone Cannot Guarantee Ethical AI. *Nature Machine Intelligence*, 3(2), 104-110.

OCSE (2021). *OECD Employment Outlook 2021: AI and the Future of Work*. OECD Publishing. <https://doi.org/10.1787/5a700c4b-en>

OECD (2023). *AI and Digital Transformation: Governance and Responsible AI*.

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

Russell, S. (2022). *Human Compatible: Artificial Intelligence and the Problem of Control*. Penguin Books.

Smith, A. (2023). *Regulatory Frameworks for a Human-Centric AI*. Routledge.

TECH4FUTURE. (2022). *Impronta di carbonio e intelligenza artificiale*. <https://tech4future.info/impronta-di-carbonio-intelligenza-artificiale/> (consultato il 9 novembre 2024).

UNESCO. (2022). *Artificial Intelligence in Education: Challenges and Opportunities*.

Whittlestone, J., Nyrupe, R., Alexandrova, A., & Dihal, K. (2021). The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions. *AI & Society*, 36(1), 1-12.

Zeng, Y., Lu, E., & Huangfu, C. (2021). Linking Artificial Intelligence Principles: The Study of Ethics Guidelines, Policy, and Practices. *Journal of Business Ethics*, 167(4), 645-662.

Zuboff, S. (2019). *Il capitalismo della sorveglianza: Il futuro dell'umanità nell'era dei nuovi poteri* (P. Bassotti, Trad.). Roma: Luiss University Press.

2. BRIDGING EXPECTATIONS AND REALITIES: THE FUTURE SOCIO-ECONOMIC IMPACT OF AI

by Viviana Condorelli*, Fiorenza Beluzzi**

Abstract: Questo studio esamina la disconnessione tra le aspettative sociali e le strategie di investimento aziendale nell'intelligenza artificiale per comprendere in che modo le scelte economiche e tecnologiche siano orientate a soddisfare i bisogni della popolazione. Attraverso l'analisi di indagini internazionali e di recenti report, esploriamo se gli investimenti aziendali rispondano ai bisogni sociali e ribadiamo la necessità di un cambiamento di prospettiva verso un'intelligenza artificiale centrata sull'uomo.

Parole chiave: IA antropocentrica, Aspettative sociali, IA generativa, Investimenti sull'IA

Abstract: This study examines the disconnection between social expectations and corporate investment strategies in AI to understand how economic and technological choices are oriented to meet the population's needs. Through the analysis of international surveys and recent reports, we explore whether corporate investments respond to social needs and propose the need for a change of perspective towards human-centred AI.

Keywords: Human-centred AI, Social expectations, Generative AI, AI investments

* Dipartimento di Fisica e Astronomia "Ettore Majorana", Università di Catania, ORCID 0009-0009-5764-2235, viviana.condorelli@phd.unict.it.

** Ricercatrice indipendente, ORCID 0009-0003-8215-685X, fiorenza.beluzzi@gmail.com.

Accettato Dicembre 2024 - Pubblicato Aprile 2025.

Introduction

The spread of artificial intelligence (AI) is transforming various areas of daily life, from healthcare to education, from entertainment to work. This technology, which has become increasingly pervasive, raises crucial questions about how it is perceived by society and how much companies' investment choices reflect social needs. According to recent surveys, the public recognizes the potential of AI to improve the quality of life, but also faces significant concerns, such as the risk of job losses and the perception of a lack of transparency in corporate decision-making processes. This study examines the disconnection between social expectations and corporate investment strategies in AI to understand how economic and technological choices are oriented to meet the population's needs. Through the analysis of international surveys and recent reports, we explore whether corporate investments respond to social needs and propose the need for a change of perspective towards human-centred AI.

1. Ethical-moral clusters

At least five clusters of ethical-moral issues that require in-depth reflection by the scientific community, policymakers, and civil society can be grouped:

- The development of AI models entails high energy and financial costs, significantly impacting energy consumption and the climate crisis, especially when energy does not come from sustainable sources. It raises a debate about the equitable distribution of AI benefits and the risk of widening the economic and technological gap between World regions.
- The datasets used often reflect a Westernized vision,

perpetuating racial, gender, and socio-economic prejudices. The emergence of monopolies in the data and AI sector, with a few companies dominating the market, accentuates these inequalities. It is, therefore, necessary to rethink management models and the role of the public sector as a guarantor of individual rights.

- The growing difficulty in distinguishing AI-generated content leads to the “automation bias”, where individuals consider AI outputs as objective and reliable, even when they are not. This increases the risk of manipulation and cognitive dependence, reducing people’s critical sense.
- Generative AI represents a challenge for cognitive jobs, with potential social tensions related to replacing or integrating human skills. According to Ipsos data, only 29% of Italian respondents believe that AI will improve the labour market, while 37% fear a worsening, indicating a negative perception of the impact of AI on employment (Ipsos, 2024).
- The need for more transparency in data collection and use raises concerns about the privacy and security of personal information. Without rigorous control, data could be exploited without individuals’ consent. Despite these risks, 47% of global respondents trust companies that use AI to protect their data (Ipsos, 2024).

2. Comparative analysis

Between 2023 and 2024, significant developments in AI have occurred, highlighting a divergence between economic investment choices and users’ perceptions, use, fears, and hopes. The AI Indexes (Maslej *et al.*, 2023; Maslej *et al.*, 2024) show that industry continues to dominate AI research, with an increasing

number of large and expensive models: in 2023, 51 significant models were produced by industry, compared to previous years (Maslej *et al.*, 2024). Training costs have skyrocketed, with models like OpenAI's GPT-4 requiring an estimated \$78 million investment (Maslej *et al.*, 2024). Additionally, AI patents have increased exponentially, especially in China, which accounted for 61.1% of AI patent origins in 2022 (Maslej *et al.*, 2024). Despite these significant investments, user awareness and understanding of AI have progressed at different rates. According to Ipsos 2023, only 50% of people knew which products and services used AI, which modestly increased to 52% in 2024 (Ipsos, 2024). This indicates a persistent gap between technological advances and public understanding of AI applications. While the percentage of individuals who say they have a good understanding of AI has remained stable at 66-67% between 2023 and 2024 (Ipsos, 2023; Ipsos, 2024), this self-assessment may not reflect actual knowledge, given the lack of awareness of practical applications.

User perceptions have evolved, showing an increase in nervousness and ethical concerns. Ipsos 2023 reported that 52% of respondents felt nervous about AI, a sentiment that remained at 50% in 2024 (Ipsos, 2024), indicating lingering concern. Trust issues remain, as 41% did not trust that companies using AI would protect personal data in 2024 (Ipsos, 2024). Additionally, incidents of AI misuse have increased, with a 32.3% increase in reported cases in 2023 (Maslej *et al.*, 2024), further fueling public concerns. Economic choices have focused on technical advances, such as developing more powerful models and increasing AI capabilities (Maslej *et al.*, 2023; Maslej *et al.*, 2024). However, users are concerned about ethical implications, including potential job displacement and discrimination. In 2024, 36% expected AI to replace their jobs, with concerns highest among younger generations (Ipsos, 2024). This contrasts with the industry's

emphasis on efficiency and performance, highlighting a misalignment between investment priorities and user concerns. Expectations that AI will profoundly change daily life remain high, with 66% anticipating a significant impact in the next 3-5 years (Ipsos, 2023; Ipsos, 2024). However, only 50% felt that AI had already changed their lives by 2024, indicating a lag between expectations and realized benefits. Users are sceptical about the economic benefits of AI; only 34% believe it will improve the economy, and 32% think it will improve the labour market (Ipsos, 2024). This scepticism highlights a disconnect between the transformative potential of AI expected by investors and the public's confidence in its benefits.

Conclusion

The evolution from 2023 to 2024 highlights a growing gap between the rapid progression of AI technologies and public perceptions and concerns. To bridge this gap, putting people at the centre of AI-related decisions is imperative, ensuring that investments improve technical capabilities and address ethical considerations and social impacts. This holistic approach balances expectations and actual investment performance in a green and integrated vision. A human-centric approach implies considering not only the performance and profitability of technology but also human well-being, dignity, and individual rights. However, the current economic context ignores these priorities in favour of profit logic. AI regulation should promote greater transparency and citizen participation in decision-making processes to ensure that individuals can understand how AI affects their lives.

In conclusion, the evolution of AI highlights a growing gap between economic logic and population concerns. To address AI's

challenges responsibly, greater transparency, open dialogue, and citizen participation are necessary. Only through a multidisciplinary approach, including social sciences, ethics, and technology, can we ensure that AI respects and promotes human values, reducing inequality and protecting individual rights.

References

Ipsos (2023, July). Global views on A.I. 2023.

Ipsos (2024, June). Ipsos AI Monitor 2024.

Maslej, N., Fattorini, L., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Ngo, H., Niebles, J. C., Parli, V., Shoham, Y., Wald, R., Clark, J., & Perrault, R. (2023, April). *The AI Index 2023 Annual Report*. AI Index Steering Committee, Institute for Human-Centered AI, Stanford University.

Maslej, N., Fattorini, L., Perrault, R., Parli, V., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J. C., Shoham, Y., Wald, R., & Clark, J. (2024, April). *The AI Index 2024 Annual Report*. AI Index Steering Committee, Institute for Human-Centered AI, Stanford University.

3. BALANCING INNOVATION AND EQUITY: AN ANALYSIS OF THE EUROPEAN AI ACT

by Sergio Pappagallo*

Abstract: Questa rubrica illustra la logica regolatoria dell'AI Act europeo (Regulation (EU) 2024/1689), fornendone un'introduzione sistematica attraverso il framework del principio di precauzione. L'analisi muove dalla definizione di intelligenza artificiale contenuta nell'Articolo 3, evidenziando le criticità tecniche e filosofiche che caratterizzano tali sistemi: dai problemi di overfitting e data bias al problema epistemologico dell'induzione nelle inferenze predittive. La rubrica dimostra come il legislatore europeo abbia strutturato una risposta normativa all'incertezza tecnologica mediante un sistema di classificazione del rischio articolato su quattro livelli progressivi: dalle pratiche espressamente proibite ai sistemi ad alto rischio, fino a quelli a rischio limitato e minimo, ciascuno soggetto a obblighi e controlli differenziati.

Parole chiave: AI Act, Innovazione, Classificazione dei rischi, Principio di Precauzione, Problema dell'Induzione

Abstract: This section illustrates the regulatory logic of the European AI Act, Regulation (EU) 2024/1689, providing a systematic introduction through the precautionary principle framework. The analysis begins with the artificial intelligence definition contained in Article 3, highlighting the technical and philosophical criticalities characterizing such systems: from overfitting and data bias problems to the epistemological problem of induction in predictive inferences. The section demonstrates how the

* PhD Student, Dipartimento di Scienze Umane, Università degli Studi "Link", <https://orcid.org/0009-0005-5960-636X>, ser.pappagallo@gmail.com.

Accettato Dicembre 2024 - Pubblicato Aprile 2025.

European legislator structured a normative response to technological uncertainty through a risk classification system articulated on four progressive levels: from expressly prohibited practices to high-risk systems, down to limited-risk and minimal-risk systems, each subject to differentiated obligations and controls.

Keywords: AI Act, Innovation, Risk Classification, Precautionary Principle, Problem of Induction

Introduction

As with most legislative processes, one of the conditions necessary to regulate a phenomenon is that the phenomenon poses risks to the values that societies need to protect, and that the phenomenon is scalable. Think of an example of the first traffic codes developed precisely because of the massive use of automobiles. Similarly, AI has been implemented for a long time already. Still, recent applications and even more recent massive uses of such technologies have made it imperative to have an initial regulation that would bring multiple actors into an international agreement. Besides, the most serious errors in AI systems require specialized skills to be identified and corrected (Binns, 2018). Therefore, it is better to anticipate risk factors. The Artificial Intelligence Act (European Commission, 2024) represents a significant attempt to balance the promotion of innovation with the need to structure the safe development and use of AI, overcoming the practices of ethics washing and protecting society from the most significant risks. This approach reflects the “responsible innovation” concept discussed by Stilgoe *et al.* (2013), which emphasizes the importance of considering new technologies’ ethical and social implications during the development process.

1. Conceptual Framework

The AI Act provides in Article 3 a definition of AI that emphasizes adaptivity post-implementation, impacting society. In addition, the definition considers the relationship between input and output. This implies that risk management must refer to this relationship. At a technical level, there are two main problems: overfitting and data bias. Overfitting occurs when a model overfits training data, compromising its ability to generalise to new data (Barocas *et al.*, 2019). Data bias emerges when the dataset does not adequately represent reality, causing potential discrimination and perpetuating existing societal biases (O’Neil, 2016). Also, note that AIs are models that use inductive inferences to make predictions. Therefore, they are subject to the classic philosophical problem of induction because even if the input data were of high quality, inductive inferences could not guarantee true conclusions by their very nature (Henderson, 2022). Therefore, uncertainty is an ineradicable component when it is necessary to obtain an amplifying prediction from the observed phenomenon to the unobserved phenomenon. Still, it is essential to govern it to ensure its manageability on a large scale. To this end, the precautionary criterion is one of the most fitting strategies in the ethical-legal context. The regulatory framework adopted by the EU on AI uses this criterion precisely, imposing preventive measures that are all the more stringent, the more significant the impact and likelihood of risks. The inherent difficulty in accurately predicting all the consequences of the massive use of AI makes the precautionary principle particularly suitable, as it promotes rigorous assessment of AI technologies’ risks and long-term implications.

2. Risk Classification

Title II of the AI Act establishes four categories of prohibited practices in using AI. The first concerns subliminal manipulation and deception: using techniques that operate beyond the users' awareness, impairing their ability to make rational choices, is prohibited. The second category concerns the exploitation of vulnerabilities: AI cannot be used to take advantage of conditions such as age, disability, or disadvantaged socio-economic situations. The third prohibition relates to discriminatory social scoring: using systems that score individuals based on social behaviors or personal characteristics is not permitted, as this reduces the person to a mere numerical datum, undermining human dignity and potentially perpetuating social inequalities. Lastly, the fourth category concerns mass surveillance and biometric identification: using real-time facial recognition and biometric tracking systems in public spaces is severely restricted and requires extreme justification and strict oversight. Then, the AI Act defines high-risk AI systems according to two key criteria. The first identifies systems that serve as security components in products already regulated by the EU in Annex II, such as medical devices, vehicles, and toys, which are subject to external compliance evaluation. The second includes systems listed in Annex III of the AI Act, including biometric identification, critical infrastructure management, education, employment, and law enforcement. The legislation provides exemptions for systems with limited impact, such as those that perform circumscribed procedural operations or optimize pre-existing human decisions. These exemptions do not include systems that profile individuals based on job performance, financial status, or behavior.

Next, the legislation defines the limited-risk categories, focusing on transparency obligations. Such technologies typically

interface with users, generate content, recognize emotions, and categorize biometric data. The legislation imposes disclosure requirements for these systems on the artificial nature of the interaction or content generated.

Finally, the minimal-risk category, which represents the majority of AI applications in the market, includes systems such as spam filters, video games, and editing tools, which are subject to minimal regulation. Although the latter do not pose significant risks to fundamental rights, the regulatory framework encourages voluntary adherence to codes of conduct to ensure ethical and transparent practices in technical implementation.

References

Acemoglu, D. (2024). *Redesigning AI: Work, Democracy, and Justice in the Age of Automation*. Cambridge, MA: MIT Press.

Acemoglu, D., & Restrepo, P. (2019). Automation and New Tasks: How Technology Displaces and Reinstates Labor. *Journal of Economic Perspectives*, 33(2), 3-30.

Albanesi, S., Marinescu, I., & Schaumann, D. (2023). The Impact of AI on Labor Markets: Evidence from Online Job Postings. National Bureau of Economic Research Working Paper Series.

Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and Machine Learning. In *Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency*.

BBC. (2023, April). We'll need universal basic income - AI's godfather'. BBC News.

Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. In *Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency*, 149-158.

Briggs, J., & Kodnani, D. (2023). The Potentially Large Effects of Artificial Intelligence on Economic Growth. Goldman Sachs Global Investment Research.

European Commission. (2022, June 16). A European approach to micro-credentials. Education and Training.

European Commission. (2024). AI Act: Regulation on artificial intelligence. Official Journal of the European Union.

Henderson, L. (Winter, 2022). The Problem of Induction. *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta & Uri Nodelman (eds.), <https://plato.stanford.edu/archives/win2022/entries/induction-problem/>.

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

Smeds, E., & Smeds, E. (2020). The Greener State: Public services for a carbon-neutral Europe. <https://core.ac.uk/download/481470986.pdf>.

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, 42(9), 1568-158.

Verlaet, T., de Kruijk, M., Rosenkranz, S., Groot, L., & Sanders, M.

(2020). Onderzoek Weten wat werkt: samen werken aan een betere bijstand, Eindrapport (What Works: Final Report). Utrecht: Utrecht University. <https://www.uu.nl/en/publication/final-report-what-works-weten-wat-werkt>.

4. LA PROTEZIONE DEI DATI PERSONALI, L'INTELLIGENZA ARTIFICIALE E I TRADUTTORI AUTOMATICI

di Federica De Stefani*

Abstract: L'intelligenza artificiale può apparentemente rappresentare una minaccia per la figura dei traduttori professionisti, ma le insidie, a ben vedere, riguardano altri aspetti, come la delicata questione del trattamento dei dati personali. Se quindi il traduttore umano, a determinate condizioni, può utilizzare l'AI per la traduzione, questo impiego deve avvenire attraverso la valutazione dei dati personali contenuti nel testo e delle specifiche modalità di trattamento che vengono garantite da ciascuno strumento.

Parole chiave: intelligenza artificiale, traduzione, traduzione automatica, protezione dei dati, traduttori professionisti

Abstract: Artificial intelligence may seem to threaten professional translators, but the pitfalls, on closer examination, concern other aspects, such as the delicate issue of personal data processing. Thus, if the human translator can use AI for translation under certain conditions, this should be done by assessing the personal data contained in the text and the specific processing methods that are guaranteed by each tool.

Keywords: Artificial intelligence, translation, automatic translation, human translator, data protection.

* Dipartimento di Scienze Umane, Università degli Studi "Link",
f.destafani@unilink.it.

Accettato Dicembre 2024 – Pubblicato Aprile 2025.

Introduzione

L'intelligenza artificiale è, ormai, molto diffusa nel campo delle traduzioni. Da Google translate a Reverso Traduzione, passando per DeepL, gli strumenti di traduzione "automatica" ai quali ci si affida per tradurre testi da una lingua all'altra stanno diventando sempre più numerosi e più utilizzati, soprattutto nelle loro versioni gratuite che, solo in alcuni casi, hanno come unico limite un numero massimo giornaliero di caratteri da poter tradurre. L'utilizzo crescente di questi sistemi si scontra, tuttavia, con una problematica che non viene quasi mai analizzata e approfondita, ossia la tutela dei dati personali¹ che viene offerta da questi strumenti. In prima battuta si potrebbe pensare che la protezione dei dati non coinvolga le operazioni di traduzione automatica, ma se si analizzano i documenti che vengono fatti tradurre da questi sistemi la realtà è ben diversa, poiché i testi spesso contengono dati personali². La questione in sé non sarebbe particolarmente complessa se il trattamento fosse accompagnato da informazioni chiare e trasparenti³, ma andando ad analizzare le condizioni di utilizzo e le informazioni rese dai vari sistemi si scopre che così non è. Le informazioni sono carenti, poco chiare e a volte persino contraddittorie⁴.

¹ G.M. Riccio, G. Scorza, E. Belisario (a cura di), *GDPR e Normativa Privacy. Commentario*, IPSOA, Milano, 2018; G. Finocchiaro (a cura di), *Il nuovo Regolamento europeo sulla privacy e sulla protezione dei dati personali*, Bologna, 2017.

² Colapietro C, Moretti A. *L'Intelligenza Artificiale nel dettato costituzionale: opportunità, incertezze e tutela dei dati personali*. BioLaw [Internet]. 2 novembre 2020.

³ In tal senso l'art. 12 del Regolamento UE 2016/679.

⁴ Sono stati analizzati oltre 20 sistemi di traduzione automatica (Word Reference, Reverso Context, Memoq, DeepL, Smartcat, Systran, Pons, Google translate, Microsoft Traduttore, iTranslate, Yandex Translate, Sonix.ai, Bing traduttore, HeyGen, Text Cortex, Phrase TMS, Lokalise, Redokun, Taia, Miraitranslate.com, modernMT): nessuno di questi affronta il tema del trattamento dei dati personali

1. Il punto dal quale partire: possiamo immettere dati personali nei sistemi di traduzione?

Il primo elemento da considerare è che non esiste un divieto assoluto di immissione dei dati personali nei sistemi di traduzione. Questo significa che la possibilità di inserire in questi sistemi i dati personali dipende dai possibili impieghi che può fare chi utilizza i traduttori automatici. Se, per esempio, il testo da tradurre, e le informazioni in esso contenute, sono protette da un vincolo di riservatezza, l'utilizzo del traduttore automatico potrebbe non rivelarsi una buona scelta. Bisogna infatti considerare che il vincolo in questione impone il divieto di comunicare a terzi le informazioni contenute nel documento da tradurre e la immissione in un traduttore automatico è, senza dubbio, una comunicazione a terzi, con annessa perdita di controllo delle informazioni. Come vengono trattati quei dati? Dove vanno a finire? Per quali scopi vengono utilizzati dall'algoritmo che governa il traduttore automatico?

Tutte queste domande dovrebbero trovare risposta nelle condizioni di utilizzo, ma, come anticipato, non sempre è così. Ci sono sistemi che forniscono informazioni contraddittorie, come Systran che nelle proprie condizioni d'uso⁵ prima dichiara che i testi tradotti con il servizio non vengono archiviati in modo permanente e non vengono condivisi né con i clienti né con terze parti, per poi precisare, nel paragrafo successivo, che i testi e le traduzioni potrebbero essere archiviati per un periodo di tempo limitato per scopi di miglioramento dell'algoritmo. Risulta

contenuti nei testi in maniera chiara e trasparente. Esistono ipotesi di divieto di utilizzo del sistema per la traduzione di testi contenenti dati personali, ma non è previsto alcun sistema di gestione degli stessi con riferimento alla protezione, all'utilizzo, alla conservazione degli stessi, sia con riguardo alle modalità sia per quanto attiene alla data retention.

⁵ <https://www.systransoft.com/systran/privacy-policy/>.

singolare, poi, la precisazione che i testi inviati e le traduzioni registrate non siano in alcun modo collegati a informazioni personali, come un indirizzo IP, quindi il servizio fornisce una indicazione relativa al fatto che i dati personali di chi utilizza il servizio non sono associati ai testi immessi e alle relative traduzioni (a differenza, per esempio di Word Reference⁶ che non solo archivia i testi immessi dagli utenti per la traduzione, ma li associa all'utente e memorizza persino la cronologia delle ricerche effettuate dall'utente) ma non c'è alcun riferimento ai dati personali eventualmente contenuti nel testo. La contraddizione, poi, continua laddove si precisa che i testi inviati ai servizi dopo aver effettuato l'accesso con un account non vengono memorizzati e vengono utilizzati solo per lo scopo necessario alla generazione della traduzione e alla fornitura del servizio, quando un paio di paragrafi prima si era specificata la possibilità di memorizzare i testi, al fine di migliorare l'algoritmo. In altri casi, come per esempio accade con DeepL⁷, si specifica che i dati vengono memorizzati per il legittimo interesse del fornitore del servizio al miglioramento dell'algoritmo e, quindi, del servizio offerto agli utenti. In altre parole, si giustifica l'utilizzo dei dati contenuti nei testi, al fine di implementare le capacità del sistema e ridurre gli errori. L'indicazione è, a parere di chi scrive, del tutto fuorviante, posto che l'utente è indotto a pensare che l'utilizzo dei dati contenuti nei testi che, vale la pena sottolinearlo, potrebbero essere dati riferiti all'utente, ma anche dati riferiti a terzi soggetti, sia necessario per avere un servizio migliore. Tecnicamente non è errato, ma il sacrificio relativo alla perdita di controllo su questi dati e al loro impiego, deve essere valutato diversamente, soprattutto quando si tratta di dati di terzi. Ci si domanda, quindi, cosa si dovrebbe fare in situazioni di questo genere. La risposta

⁶ <https://www.wordreference.com/english/TermsOfService.aspx>.

⁷ Si veda <https://www.deepl.com/it/privacy>.

appare tanto banale quanto scontata: si dovrebbe evitare l'utilizzo dello strumento. La scelta, tuttavia, deve basarsi sull'analisi, oggettiva, delle garanzie e delle condizioni che vengono imposte dal servizio e questo impone, ovviamente, la lettura e l'analisi delle condizioni di utilizzo.

2. Quali sono le garanzie che dovrebbero essere offerte?

Considerando l'iter che deve portare un soggetto all'utilizzo del traduttore automatico, il primo aspetto da valutare è relativo all'eventuale accordo di riservatezza che dovesse riguardare il testo da tradurre. Laddove fosse presente un simile patto è evidente che le scelte, obbligate, sarebbero solo due: desistere dall'utilizzo o utilizzare sistemi che forniscano adeguate garanzie in merito al trattamento dei dati. Questa seconda possibilità si traduce non solo nella analisi, come detto, delle caratteristiche tecniche del sistema di traduzione (che non saranno mai complete (Hosni, Vulpiani, 2023) trattandosi di sistemi algoritmici che presentano in ogni caso delle black box), ma anche nella sottoscrizione di eventuali accordi specifici con il fornitore del servizio. Ci sono infatti sistemi che espressamente escludono la possibilità di utilizzare il servizio per scopi lavorativi o professionali, ma anche questo aspetto non sempre è chiaro poiché, di base, manca la lettura delle condizioni di utilizzo da parte degli utenti, ma sarebbe opportuno delimitare utilizzi e impieghi dei dati con un accordo ad hoc. Le difficoltà di trovare un accordo con il fornitore per l'utilizzo di un sistema di traduzione sono evidenti, ma la delimitazione degli impieghi e degli utilizzi dei dati è fondamentale per poter utilizzare il sistema di traduzione automatica perimetrando le rispettive, eventuali, responsabilità.

3. Quali sono le criticità in termini di responsabilità?

Le criticità attinenti all'impiego delle traduzioni automatizzate si possono dividere in due diverse categorie: la responsabilità per la comunicazione non autorizzata dei dati a terzi e la responsabilità per eventuali ulteriori utilizzi dei dati effettuati da soggetti terzi.

Rientra nella prima categoria l'impiego di sistemi di traduzione che memorizzano i dati laddove il contenuto del testo tradotto sia soggetto ad un vincolo di riservatezza. In questo caso, la comunicazione a terzi è vietata e l'impiego di traduttori automatici è da escludere. Sarebbe altresì opportuno specificare per iscritto tale divieto, proprio per evitare spiacevoli inconvenienti dovuti all'impiego delle informazioni riservate da parte dell'algoritmo di traduzione. Rientrano, invece, nella seconda ipotesi i casi in cui il traduttore automatico sia ammesso dal rapporto contrattuale che lega l'utente al soggetto che fornisce il testo da tradurre, ma a condizione che di questi dati non ne venga fatto alcun diverso impiego. In questo caso, la scelta del traduttore da utilizzare diventa fondamentale poiché le caratteristiche tecniche del sistema, eventualmente associate ad un accordo tra utente e fornitore, perimetrano e delineano le rispettive responsabilità in caso di utilizzo dei dati.

4. Conclusioni

L'impiego dei sistemi di traduzione automatica è ormai molto diffuso in tutti i settori, ma tale diffusione non si accompagna ad una adeguata conoscenza dello strumento che si utilizza. L'attenzione, come anticipato, è focalizzata sull'output prodotto dal sistema di traduzione, sulla sua immediatezza e, in

moltissimi casi, sulla gratuità del servizio. Questo tipo di approccio fa passare in secondo piano il vero problema collegato all'impiego di questi sistemi, ossia la reale ed efficace protezione dei dati che vengono immessi per la traduzione. Come è emerso dall'analisi svolta sul campione di traduttori automatici le informazioni che vengono fornite agli utenti sono poco chiare e persino contraddittorie e ci si domanda, quindi, quali possano essere le soluzioni per questo status quo che appare difficile da sovvertire e modificare. Il primo, indispensabile elemento dal quale partire è la conoscenza dei sistemi di traduzione che si utilizzano, conoscenza finalizzata all'acquisizione della consapevolezza che ogni sistema può avere caratteristiche diverse e può, conseguentemente, trattare e proteggere i dati con modalità differenti. Tale presa di coscienza consente all'utente di scegliere non solo sistemi di traduzione che offrano idonee garanzie, ma anche di valutare quale tipologia di dati immettere in questi sistemi. I rischi che possono presentare questi strumenti sono molteplici e sono legati alle vulnerabilità che possono derivare dall'adozione di sistemi che non siano conformi alla normativa. Basti pensare, a tal riguardo, alla possibile archiviazione in cloud non compliant che espongono i dati al rischio di accessi abusivi e a possibili violazioni degli stessi, oltre, ovviamente, all'utilizzo per l'addestramento degli algoritmi, con conseguente perdita di qualsivoglia controllo sugli stessi. Se da un lato i rischi sono chiari, dall'altro le soluzioni che potrebbero mitigare tali criticità possono essere racchiuse in due diverse categorie, una formativa e l'altra tecnica. La formazione è senza dubbio un elemento che risulta imprescindibile, poiché solo attraverso la conoscenza dello strumento che si utilizza, del suo funzionamento, delle sue caratteristiche, delle conseguenze che possono derivare da un suo impiego si può mettere l'utente nelle condizioni di effettuare scelte consapevoli. Dall'altro lato è fondamentale considerare le garanzie tecniche offerte da questi

sistemi, magari attraverso certificazioni e conformità a normative internazionali sulla sicurezza informatica. Solo attraverso queste due diverse impostazioni, la valutazione tecnica e la consapevolezza dei rischi connessi all'impiego di tali sistemi l'utente può effettuare scelte coscienti relative non solo all'eventuale impiego del traduttore automatico, ma anche relative alla tipologia di traduttore, scegliendolo tra i diversi sistemi che il mercato offre, oltre a individuare quali dati e quali informazioni utilizzare per la traduzione automatica e, quindi, immettere nel sistema.

Bibliografia

Colapietro, C., & Moretti, A. (2020). L'Intelligenza Artificiale nel dettato costituzionale: opportunità, incertezze e tutela dei dati personali. *BioLaw Journal - Rivista Di BioDiritto*, 3, 359–387. <https://doi.org/10.15168/2284-4503-713>.

Finocchiaro, G., (a cura di). (2017). *Il nuovo Regolamento europeo sulla privacy e sulla protezione dei dati personali*. Bologna: Zanichelli.

Hosni, H., Vulpiani, A. (2023). Algoritmi e previsioni. *Sistemi intelligenti, Rivista quadrimestrale di scienze cognitive e di intelligenza artificiale*, 1, 143-156.

Regolamento UE 2016/679 del Parlamento europeo e del Consiglio del 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE

(regolamento generale sulla protezione dei dati). <https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32016 R0679>.

Riccio, G.M., Scorza, G., Belisario, E., (a cura di). (2018), *GDPR e Normativa Privacy. Commentario*, IPSOA, Milano: Wolters Kluwer.

Sitografia

<https://www.systransoft.com/systran/privacy-policy/>.

<https://www.wordreference.com/english/TermsOfService.aspx>.

<https://www.deepl.com/it/privacy>.

5. PARTICIPATORY APPROACHES FOR THE TRANSITION FROM AUTOMATION TO ARTIFICIAL INTELLIGENCE (AI): A CASE STUDY

by Sara Calicchia, Chiara Colagiacomo, Angela Bagnato, Roberta Pistagni, Bruno Papaleo, Francesca Grosso*

Abstract: Per bilanciare il progresso tecnologico con il benessere dei lavoratori è fondamentale preservare spazi di partecipazione centrati sull'uomo e integrare i programmi tradizionali di sicurezza con iniziative di promozione della salute che rafforzino il ruolo attivo dei lavoratori. Questo studio presenta il caso di un workshop partecipativo, condotto all'interno di un'azienda italiana del settore automotive, durante il quale lavoratori, manager e esperti si sono confrontati su come migliorare le condizioni di lavoro, anche attraverso l'uso di sistemi automatizzati di intelligenza artificiale (IA). Attraverso la mappatura dell'esperienza e il dialogo strutturato, sono emerse soluzioni concrete per ottimizzare i processi e promuovere il benessere. I risultati evidenziano un

* Sara Calicchia, INAIL, Dipartimento di medicina, epidemiologia, igiene del lavoro e ambientale, Istituto Nazionale per l'Assicurazione contro gli infortuni sul lavoro, <https://orcid.org/0000-0003-2171-331X>, s.calicchia@inail.it. Chiara Colagiacomo, INAIL, Dipartimento di medicina, epidemiologia, igiene del lavoro e ambientale, Istituto Nazionale per l'Assicurazione contro gli infortuni sul lavoro, c.colagiacomo@inail.it. Angela Bagnato, Unitelma Sapienza Università di Roma, angela.bagnato@unitelmasapienza.it. Roberta Pistagni, INAIL, Dipartimento di medicina, epidemiologia, igiene del lavoro e ambientale, Istituto Nazionale per l'Assicurazione contro gli infortuni sul lavoro, r.pistagni@inail.it. Bruno Papaleo, INAIL, Dipartimento di medicina, epidemiologia, igiene del lavoro e ambientale, Istituto Nazionale per l'Assicurazione contro gli infortuni sul lavoro, b.papaleo@inail.it. Francesca Grosso, INAIL, Dipartimento di medicina, epidemiologia, igiene del lavoro e ambientale, Istituto Nazionale per l'Assicurazione contro gli infortuni sul lavoro, f.grosso@inail.it.

Accettato Dicembre 2024 - Pubblicato Aprile 2025.

miglioramento dell'organizzazione del lavoro e delle relazioni interpersonali, rafforzando la fiducia nel cambiamento tecnologico. L'approccio partecipativo si conferma strategico per integrare le tecnologie IA in modo equo e inclusivo, migliorando la capacità decisionale e il coinvolgimento dei lavoratori.

Parole chiave: intelligenza artificiale, approccio partecipativo, promozione della salute, sviluppo organizzativo, approccio centrato sull'uomo.

Abstract: To balance technological progress with worker well-being, it is essential to maintain human-centered spaces for participation and integrate traditional safety programs with health promotion initiatives that empower workers. This study presents a participatory workshop conducted in an Italian automotive company, where workers, managers, and experts collaborated to improve working conditions, taking into consideration AI-based solutions as well. Through experience mapping and structured dialogue, concrete solutions emerged to optimize processes and enhance well-being. The results indicate improvements in work organization and interpersonal relationships, fostering greater trust in technological change. A participatory approach is a key strategy for integrating AI technologies fairly and inclusively, enhancing decision-making and worker engagement.

Keywords: artificial intelligence, participatory approach, health promotion, organizational development, human-centered approach.

Introduction

Artificial Intelligence (AI) is expected to profoundly influence key dimensions affecting worker well-being in the coming years, such as autonomy, interpersonal relationships, task integrity, and skill utilization (Bankins *et al.*, 2023). In the specific field of occupational health and safety, introducing AI-based

worker monitoring and control technologies, such as self-tracking devices, risks further shifting safety responsibilities onto individual workers. These systems often focus solely on risks identified by AI, neglecting organizational risks that are not technologically monitored (Malenfer *et al.*, 2023). This may mark a return to the Cartesian paradigm of human reductionism and the biomedical approach (Till, 2018), creating new forms of worker control and fostering distrust toward new technologies. To harness the potential of emerging technologies without compromising individual and organizational well-being, it is essential to preserve or restore spaces for participation and reflection, unmediated by algorithms, that prioritize human needs. In Occupational Health and Safety (OHS), it is particularly urgent to integrate traditional risk assessment and prevention programs with Workplace Health Promotion (WHP) initiatives that empower workers to influence the social determinants of their health. This approach aligns with the Luxembourg Declaration and recent comprehensive frameworks for health protection and promotion (Pronk, 2013).

1. Objectives and Methodology

This case study, conducted within a leading automotive company in Italy, introduces the methodology and key outcomes of a participatory workshop involving workers and managers, supported by experts in sustainable organizational models, facilitation techniques, and process engineering. The workshop aimed to improve working conditions in production cells following an assessment process that involved most of the company's personnel (Pistagni *et al.*, 2023). The goal was to identify solutions to alleviate workload in specific production lines through greater automation. The initiative, carried out alongside the

implementation of an AI-based Worker Management (AIWM) project, sought to enhance worker participation and exemplify best practices presented by the company in various institutional settings. The pilot workshop focused on a specific work cell with identified improvement needs. The facilitation methodology was inspired by the “experience map” approach (Strickland *et al.*, 2019), a participatory technique that actively engages employees in analyzing and visualizing their work experiences. This approach aims to optimize processes, promote well-being, and create a more inclusive and health-conducive work environment, ultimately improving productivity and job satisfaction. Experts gathered precise insights into the production process through detailed illustrations of individual actions and phases of work mapped by workers and managers. Workers then identified critical areas on the process map, followed by structured reflection exercises to examine real-world challenges. Using written prompts, participants articulated their thoughts on sticky notes, a method that fosters clarity and critical analysis. Group discussions involving workers, managers, and experts generated actionable proposals and solutions based on these inputs.

The development process was evaluated using focus groups and surveys targeting managers and workers. The surveys captured perceptions of organizational changes, improvements in manager-worker relations, and factual data on implemented projects.

2. Results

This section briefly presents the results of the participatory process, specifically focusing on the workshop conducted in the operational cell. Out of a total of 30 workers, 11 responded to the

qualitative-quantitative questionnaire, while 15 attended the focus groups conducted at the end of the process. In line with the integrated OSH/WHP strategy, workers reported significant improvements in their work conditions, ranging from the physical environment to organizational practices. *“I feel relieved because it means that management also considers workers’ needs”*. Open discussions were among the most valued aspects of the process. The operational cell workshop was particularly appreciated for its concreteness and specificity. Participants highlighted the role of small group work in facilitating mutual understanding and deeper exploration of issues: *“Addressing the problems was more enjoyable, we understand each other better, and we’re fewer people”*; *“Compared to the first time, I found this more concrete, really aimed at solving the situation”*; *“Now we went into detail, narrowing down the focus and reaching a final solution”*.

Managers also highlighted the generative power of the experience as an opportunity to learn from those on the front lines: *“It was the most interesting thing we’ve done here: I realized, through the way they spoke in their own terms, with their feelings, some very precisely, just how important they feel about what they are doing”*. Sharing the map of the cell and working on it together yielded unexpected results due to its concreteness: *“That map is normality for me, but having it printed and laid out on the table had an unimaginable effect”*.

Additionally, the workshop identified improvement solutions for critical issues highlighted during the risk assessment phase, lending credibility to implementing AI technologies designed to address them (e.g., a visual control camera relieving workers from certain repetitive gestures). The participatory process further supported the need to optimize work areas with new production lines integrated into the company’s management system, featuring enhanced human-machine interfaces (HMIs) that simplify

operational and maintenance activities. Moreover, workflow efficiency improved by introducing advanced technologies and wearable sensors for health and safety monitoring, developed in collaboration with various research institutions.

Conclusion

The findings confirm that participatory occupational health and safety approaches foster continuous learning between managers and workers, facilitating intersectoral knowledge exchange. This framework supports the implementation of fair, inclusive AI technologies that address workers' needs. Participatory methods enhance decision-making quality by incorporating workers' practical knowledge of tasks and challenges, yielding more informed, realistic, and effective outcomes. Successful organizational changes that improve worker well-being depend on careful people management that builds a shared vision within the organization. This approach delivers impactful, context-specific solutions that optimize processes. Such participatory processes can mitigate the risk of AI being perceived or used as a control tool that infringes on privacy. Future research should explore how participatory organizational development initiatives influence worker acceptance of new technologies.

Acknowledgments

We would like to express our special thanks to Oerlikon Friction Systems (Italy) Srl for their valuable support in carrying out this project. We would also like to thank all the staff we worked with for their willingness to give us time so generously.

References

Bankins, S., Formosa, P. (2023). The ethical implications of artificial intelligence (AI) for meaningful work. *J Bus Ethics*, 185(4), 725-740. doi.org/10.1007/s10551-023-05339-7.

Malenfer, M., Sarrey, M., Clerté, J., Hery, M., *et al.* (2023). Artificial intelligence in the service of health and safety at work: Perspectives and challenges from now to 2035 - A prospective study. *Qeios*. doi:10.32388/NRENI6.

Pistagni, R., Calicchia, S., Grosso, F., Colagiacomo, C., Papaleo, P. (2023). Fare ricerca partecipata attraverso coaching e facilitazione. *QTimes – webmagazine*. Anicia Editore. 2(1), 449-463. doi: 10.14668/QTimes_15169.

Pronk, NP. (2013). Integrated worker health protection and promotion programs: overview and perspectives on health and economic outcomes. *J Occup Environ Med.*, 55(12 Suppl), S30-7. doi: 10.1097/JOM.0000000000000031. PMID: 24284747; PMCID: PMC4155035.

Strickland JR., Kinghorn AM., Evanoff BA., Dale AM. (2019). Implementation of the Healthy Workplace Participatory Program in a Retail Setting: A Feasibility Study and Framework for Evaluation. *Int J Environ Res Public Health*, 16(4), 590. doi: 10.3390/ijerph16040590. PMID: 30781669; PMCID: PMC6406806.

Till, C. (2018). Commercialising bodies: action, subjectivity and the new corporate health ethic. In: Lynch, R., Farrington, C. (eds) *Quantified Lives and Vital Data. Health, Technology and Society*. London: Palgrave Macmillan. doi.org/10.1057/978-1-349-95235-9_10.

SAGGI

1. INTELLIGENZA ARTIFICIALE GENERATIVA, COSTRUZIONE DI SENSO E RETI SOCIALI: UNA PROSPETTIVA SOCIOLOGICA

di Francesco Mattioli*

Abstract: Sul ruolo e le prospettive dell'A.I. la sociologia offre notevoli contributi teorici e di ricerca. Innanzitutto, nello studio dei processi di negoziazione e di costruzione di senso. L'A.I. sembrerebbe poter ridurre il grado di incertezza nella cosiddetta società liquida, ma potrebbe generare una complessità legata al confronto tra una logica algoritmica e certi valori etici condivisi. Inoltre, vanno citati gli studi sulle reti sociali, che sono generate da sistemi di interazione in grado di influenzare la condotta individuale e collettiva a partire dagli sviluppi più avanzati della comunicazione mediatica, oggi governata dall'A.I. Così, la società intesa come rete acquisterebbe sempre più valore euristico nell'interpretazione dei processi sociali. Infine, molti studi riguardano il ruolo dell'A.I. nel cambiamento sociale, sia nel più ampio contesto dello sviluppo sostenibile, sia a fronte della possibilità di riequilibrare le differenze tra paesi ricchi e paesi poveri, sia ancora nello studio della differenziazione sociale di origine generazionale.

Parole chiave: sociologia, costruzione di senso, algoritmica, rete, cambiamento sociale.

Abstract: On the role and prospects of AI, sociology offers significant theoretical and research contributions. First of all, in the study of the

* Già professore ordinario di Sociologia, Dipartimento di Sociologia e Comunicazione, "Sapienza" Università di Roma, francesco.mattioli@fondazione.uniroma1.it.

Accettato Dicembre 2024 - Pubblicato Aprile 2025.

processes of negotiation and construction of meaning. AI would seem to be able to reduce the degree of uncertainty in the so-called liquid society, but it could generate a complexity linked to the comparison between an algorithmic logic and certain shared ethical values. Furthermore, studies on social networks should be mentioned, which are generated by interaction systems capable of influencing individual and collective conduct, starting from the most advanced developments in media communication, today governed by AI. Thus, society, as a network, would acquire an increasingly heuristic value in the interpretation of social processes. Finally, many studies concern the role of AI in social change, both in the broader context of sustainable development, and in the face of the possibility of rebalancing the differences between rich and poor countries, and still in the study of social differentiation of generational origin.

Key words: sociology, construction of meaning, algorethics, network, social change.

1. La società dell'incertezza

Globalizzazione e postmodernità sono concetti che a vario titolo hanno segnato il firmamento del pensiero sociologico almeno a partire dagli anni '70 del secolo scorso. Quello che essi ci restituiscono non è il quadro di una società contemporanea soggetta ad un progressivo processo di standardizzazione e di omologazione, come temevano taluni critici della vecchia scuola apocalittica franco-tedesca, da Eric Fromm a Herbert Marcuse, da Jean Baudrillard ad Alain Touraine. La chiave che ha girato in altra direzione tale prospettiva è stato certamente il processo di "glocalizzazione" che, proprio in grazia dei media, ha condotto sulla scena internazionale specificità etniche, culturali e politico-sociali tali da punteggiarla di soggetti e di soggettività distinte, eterogenee,

producendo un crescente stato di pluralità e quindi di complessità.

Così, il XXI secolo si è aperto con nuovi paradigmi interpretativi che tengono conto della complessità e ne fanno anzi il punto di riferimento per l'interpretazione delle dinamiche socioeconomiche, politiche e socioculturali emergenti nel nuovo millennio. Il primo, e il più noto, è certamente quello di Zygmund Bauman (2000) che descrive una società "liquida", caratterizzata dall'incertezza, innescando un dibattito anche sul relativismo, o nelle migliori delle ipotesi sul relazionismo, per dirla con Karl Mannheim. L'altro, è quello proposto da Nassim N. Taleb (2007) con la sua "teoria del cigno nero", che sottolinea la forte casualità degli eventi e la necessità di creare modelli interpretativi della realtà affidati ad un più o meno efficace calcolo della probabilità. Modelli che suggeriscono un'idea se non proprio di dis-ordine (il concetto di entropia, valido per qualsivoglia sistema energetico, non è estraneo alla questione...) certamente di crescente eterogeneità che impegna lo scienziato sociale sia nella fase di spiegazione di un fenomeno, sia e ancor più nella previsione degli eventi. Nella spiegazione è necessario selezionare quelle variabili che in modo più evidente sembrano connesse con l'evolversi di un fenomeno sociale, trascurandone altre – meno evidenti e forse apparentemente meno determinanti – che pure ne sono anch'esse responsabili. Nella previsione, come accennato a proposito del pensiero di Taleb, dobbiamo ricorrere ad un calcolo delle probabilità che, se soccorre nell'agire pratico, non mette al riparo dal possibile errore e quindi induce ad adottare sovente una più prudente strategia di precauzione. Fino ad un certo punto, si era creduto che fosse questa la differenza tra le scienze fisiche e le scienze sociali; che mentre le prime fossero in grado di proporre delle "verità" pressoché oggettive con il conforto della logica, della matematica e di una efficace applicazione in campo tecnologico, le seconde subissero l'imprevedibilità del loro oggetto di studio,

l’Uomo, che essendo dotato di volontà propria poteva in qualsiasi momento divergere – per motivi istintivi, soggettivi, culturali, storici – dalla previsione, rendendo precario il lavoro sia del sociologo che del policy maker. Ma poi è intervenuto un punto di svolta che ha rivoluzionato la scienza nella sua interezza, rimescolandone le carte: lo sviluppo della fisica quantistica. Non è questa la sede per approfondire più di tanto il tema, ma sarà opportuno sottolineare almeno tre aspetti che qui possono interessare maggiormente. Il primo è costituito dal “principio di indeterminazione” enunciato da Werner Heisenberg nel 1929: considerando la doppia natura delle particelle elementari della realtà fisica – sia materia, che energia – non si può avere una conoscenza completa di esse, perché se ne osserviamo il moto perdiamo di vista la posizione, e viceversa, se conosciamo la posizione, non possiamo determinarne in modo accurato il moto.

Di conseguenza della realtà avremo sempre una conoscenza approssimata, semmai probabilistica e comunque legata alle nostre capacità di osservazione. Il secondo aspetto si deve al contributo di Erwin Schrödinger, il quale nel 1935 formula il principio di sovrapposizione (*entanglement*) delle particelle, che comporta da un lato il fatto che finché non si procede ad un esame empirico non si può essere sicuri dell’esistenza di una realtà o del suo contrario, e dall’altro che due particelle che si incontrano conservano memoria l’una dell’altra, creando legami indissolubili ma imprevedibili. Un terzo aspetto è rappresentato dal “principio antropico”, formulato inizialmente nel 1974 dal fisico Brandon Carter e ripreso da altri scienziati, ad esempio da J.D.Barrow, Frank J. Tipler, John A. Wheeler, secondo il quale la realtà ha senso soltanto se e in quanto “conosciuta” e “definita” da un osservatore: quindi essa esiste di fatto solo in virtù di un Essere Umano che vi entra in relazione con le proprie facoltà percettive e riflessive.

Si comprende allora come tutta la conoscenza scientifica, per quanto sostenuta dalla logica, dalla matematica, dall'esperienza empirica e dalla sperimentazione e quindi più condivisibile di altre forme conoscitive che si fondano sul mero intuito o su un ragionamento esclusivamente logico, dalla filosofia alla teologia, presenti essa stessa dei limiti. Limiti che vengono ridotti, limati e corretti mediante il ricorso ad un calcolo sempre più raffinato delle probabilità e alla continua sperimentazione, ma che non consentono di invocare l'oggettività: quella oggettività che era il cavallo di battaglia della scienza positivista e che oggi viene meno, riservata ad approcci di carattere fideistico, ideologico o religioso che qui esulano dalle nostre riflessioni. Ma c'è di più. Altri elementi intervengono a rendere complessa la lettura umana della realtà e a far traballare le certezze della scienza. Uno di questi emerge dalla discussione del concetto di entropia. Considerata come la misura del disordine, che progressivamente e inevitabilmente caratterizza il farsi dell'universo, ad essa viene contrapposto un processo inverso, di recupero d'ordine, sia attraverso l'informazione (notoriamente intesa come riduzione dell'incertezza), sia attraverso l'adattamento ambientale delle forme biologiche, sia infine in ragione della volontà creatrice che secondo Pierre Teilhard de Chardin caratterizzerebbe la "singolarità" della specie umana. Su questi temi si possono prendere in considerazione soprattutto i contributi innovativi di scienziati e di filosofi della scienza come lo stesso Schrödinger, di Luigi Fantappiè, di Ilya Prigogine, di Humberto Maturana e Francisco Varela, ma anche – seppur per altri versi – di Claude Shannon. Per una rilettura critica della scienza, inoltre, non si può dimenticare il fondamentale contributo di Karl Popper (1935/1959), il quale ha spiegato come ogni verità scientifica abbia valore solo "fino a prova contraria". Infine, va sottolineato l'apporto della sociologia della scienza. Per questa disciplina si deve partire dagli studi di Robert K. Merton (1957) e di

Thomas Kuhn (1962), secondo i quali le “verità” scientifiche seguono o comunque interagiscono con i cambiamenti storici e assumono differenti significati e differenti livelli di credibilità a seconda delle dinamiche sociali che si sviluppano entro la comunità – o meglio, le comunità – degli scienziati. In una società di cui si sono descritte la “caduta delle ideologie”, la crisi delle religioni e la fungibilità di qualsivoglia credenza offerta dal sistema dei media, accade allora che le criticità emerse anche nel sapere scientifico tra indeterminazione e probabilismo costituiscano una ulteriore spinta verso quella che Massimo Sideri, conversando con Taleb sul numero di ottobre 2020 di *Corriere Innovazione*, ha definito come “incertocene”, un’era di stabile e tangibile senso di incertezza.

2. Gen-AI

È in questo preciso punto della storia umana che interviene l’Intelligenza Artificiale (d’ora in avanti, per brevità la chiameremo A.I.).

Se ne è scritto e se ne sta scrivendo talmente tanto sulla A.I., che qui appare pressoché superfluo approfondirne genesi e modalità di crescita. È preferibile mettere alcuni puntini sulle *i*, al fine di concentrarsi su quel che può significare per la sociologia. Che la A.I. giochi un ruolo importante nel cercare di dirimere i dubbi sulla società dell’incertezza e sull’evoluzione della scienza all’ombra della fisica quantistica è indubitabile. La A.I. infatti dovrebbe aiutare a compiere scelte razionali e condivise, e a farlo in tempi rapidi, quindi con una efficienza improponibile per le facoltà dell’Uomo. Lo stesso calcolo delle probabilità, su cui sembra ormai basarsi la conoscenza umana, dovrebbe giungere a livelli di attendibilità estremamente elevati proprio in grazia della capacità

della A.I. di raccogliere e di valutare con raffinata attenzione tutte le variabili intervenienti, nella loro qualità e nel loro peso. Di fronte ad una crescita impetuosa dell'efficienza della A.I., non sarà marginale ricordare che quantunque oggi gli algoritmi che sono alla base della A.I. funzionino per lo più ancora secondo una logica binaria – quindi 0/1 – cominciano ad essere disponibili dei computer cosiddetti “quantistici” in grado di tener conto di un ventaglio di possibilità poste tra 0 e 1, ciascuna con il suo grado di probabilità.

Di classificazioni della A.I. ve ne sono a iosa. Una delle più semplici, e seguite, è quella che attualmente distingue tra *A.I. debole* e *A.I. forte*, o meglio *generativa*. La prima è quella che è stata finora di uso comune, e della quale ormai neppure ci si rende più conto, tanto è diffusa. Ha portato a progressi rilevanti nel campo della mecatronica, ad esempio nell'industria automobilistica e nel settore della robotica sanitaria, e nella sua versione più spinta riesce ad essere “predittiva” perché sulla base delle informazioni di cui dispone e degli algoritmi adatti all'uopo, ti dice cosa “succede se”. In quest'ultima veste si adotta ad esempio nelle politiche aziendali, nella finanza, persino nella meteorologia. *L'Intelligenza Artificiale Generativa* (detta anche *Gen-AI*) invece non si limita a scegliere una tra le alternative possibili precostituite, non si limita a predire, ma costruisce opzioni nuove collegando autonomamente tutte le informazioni che le vengono fornite. Quindi Gen-AI possiede tutti gli elementi per confezionare prodotti intellettuali che, meglio che nell'Uomo, sono il risultato di un'accurata analisi di tutto il sapere disponibile, per di più in brevissimo tempo. Per conseguenza sembrerebbe in grado di determinare ciò che è definibile come “vero” o “corretto” e ciò che non lo è, e nella sua versione più avanzata – quella quantistica e dunque probabilistica – saprebbe anche discernere le opportunità, perché non vedrebbe solo il bianco o il nero delle cose, ma anche le sfumature, i possibili

compromessi. La realizzazione e il continuo perfezionamento di Gen-AI, come è noto, dipende molto, oltre che dai progressi della cibernetica e dell'informatica, da quelli delle neuroscienze, che avrebbero individuato i meccanismi che stanno alla base del sistema decisionale proprio dell'Essere Umano.

Quando, nella rete neurale, il neurone riceve informazioni da altri neuroni, crea per una sua stessa natura delle selezioni, delle associazioni e delle rielaborazioni dei dati, apparentemente autonome, che poi ritrasmette in rete. Tutto ciò arricchisce, dilata, complica la produzione di informazioni che servono sia a guidare il corpo che la mente dell'individuo. Non è chiaro se tutto ciò sia sufficiente a spiegare le scelte e le decisioni umane, e quindi abbia a che fare con la soggettività e la creatività o meno, ma certo nella mente umana le reti neurali creano nuove forme di apprendimento profondo (*deep learning*) e di conoscenza e permettono al soggetto di orientarsi e di avviare comportamenti intenzionali. Il fatto che ogni singolo neurone abbia la capacità di rielaborare l'informazione che gli proviene dai neuroni con cui è collegato in rete, ha fatto pensare alla possibilità di creare una macchina funzionante allo stesso modo, quindi in grado di perfezionare progressivamente le sue operazioni e di orientarle secondo determinati fini. Gino Roncaglia (2023), a questo proposito, fa notare che Gen-AI già oggi non si limita a pescare il meglio dalle enciclopedie di dati che le vengono immesse in memoria, ma ne fa un uso originale in virtù della sua capacità di rielaborare in breve tempo tutte le informazioni di cui dispone, più e meglio dell'Essere umano. È proprio per questo, del resto, il motivo per cui l'Uomo si è dotato di Gen-AI: la possibilità di disporre di uno strumento di grande efficienza informativa e operativa per portare a buon compimento i propri obiettivi. La risposta dell'Intelligenza Artificiale Generativa allora non sarebbe automatica, scontata, come accade quando si smanetta su una calcolatrice tascabile, ma in qualche misura

autonoma, per così dire “creativa”, cioè adattiva alla bisogna. Entrerebbero in gioco, infatti, un gran numero di parametri che permetterebbero di modellizzare ad hoc, proprio come avviene nel caso umano, i più vari comportamenti di scelta. È in tale prospettiva che vanno intesi gli sforzi di riprodurre con Gen-AI i ragionamenti umani, riducendo la loro variabilità soggettiva quando occorre ottenere processi conoscitivi, interpretativi e decisionali pressoché indiscutibili e perciò stesso di grande efficacia, ma allo stesso tempo riservandosi la possibilità di mantenere una certa flessibilità, una maggiore duttilità nelle scelte, insomma una costante capacità di autocorrezione dei fini e dei mezzi. Il che ridurrebbe anche il rischio di non riuscire ad adattare la risposta a problemi che cambiassero d'improvviso il loro aspetto o la loro entità. Un esempio? Mantenere flessibilità di risposta a fenomeni idrogeologici e sismici dei quali non si possono conoscere a pieno le tempistiche o le modalità con cui si verificano. Saremmo dunque vicini alla realizzazione della Macchina di Turing, di quella, finora soltanto ipotizzata, in grado di poter rivaleggiare con l'Uomo? Secondo Turing non era possibile, ma i futurologi prospettano la possibilità di giungere ad una terza fase della A.I., una “Superintelligenza Artificiale” in grado di sopravvivere a tal punto le capacità degli esseri umani da trovarsi in posizione di vantaggio rispetto ad essi nelle decisioni. Ma in questo caso siamo più vicini alla fantascienza che alla futurologia.

Al momento, deve essere chiaro che Gen-AI non agisce da sola: perché come avviene nella progettazione di algoritmi, dove l'input è ciò che l'Uomo immette e l'output è ciò che l'Uomo chiede, quel che essa può fare dipende da quel che l'Uomo ha immesso nella memoria della Macchina e solo in base alle richieste da lui formulate. Tutto ciò sembrerebbe testimoniare che, al di là dell'allarme di chi intravede un futuro sventurato per l'Uomo, preda – come voleva intendere il matematico e scrittore Vernor

Vinge – di una “singolarità tecnologica” in grado di superare rapidamente le facoltà umane, le preoccupazioni legate allo sviluppo della A. I sono di altro tenore. Se infatti si vanno a leggere i comunicati e le dichiarazioni di principio delle maggiori organizzazioni internazionali e nazionali, dalle Nazioni Unite all’Unione Europea, dallo Stato Italiano al Vaticano, e gli stessi timori espressi da alcuni dei “padri” dell’attuale Gen-AI, come il recente Premio Nobel per la fisica Geoffrey Hinton e Sam Altman, si noterà che le cautele riguardano aspetti ben più concreti e contemporanei. Uno dei primi “scontri” sulla A.I. è avvenuto infatti quando il New York Times ha fatto causa a Microsoft e a Open AI per i miliardi di dollari di danni causati dalla copia e dall’utilizzo illegale, mediante il noto chatbot ChatGPT3, di materiale elaborato e proposto dal giornale sulle proprie pagine. A cascata ne sono venute forti rivendicazioni sul diritto d’autore in genere. Un secondo punto che si ritrova nei tanti documenti ufficiali ha a che vedere con la necessità di estendere i benefici della A.I. a tutta la popolazione del pianeta, evitando di replicare quel “digital divide” che nei decenni passati ha creato distorsioni e ritardi sociali, economici e culturali nei paesi del cosiddetto Terzo Mondo. Un altro argomento che inquieta riguarda l’applicazione della A.I. ai sistemi di comunicazione; già ora c’è la possibilità di diffondere inaffidabili fake news, improbabili *deepfakes*, di manipolare e distorcere l’informazione, creando perfino personaggi inesistenti in grado di farsi influencer di centinaia di migliaia di followers e di spingere molte persone a vivere una falsa vita nel metaverso piuttosto che in quella della realtà quotidiana. Così, il discorso si sposta sul controllo sociale, sulla democrazia e ovviamente sugli strumenti di guerra, sospingendo ciecamente le meraviglie della A.I. nelle grinfie di certo misoneismo, del più vieto complottismo e del più pericoloso avventurismo. Al che sorge la domanda se il futuro della A.I. sia *per* l’Uomo, *con* l’Uomo oppure *oltre* l’Uomo; domanda

che continuano a porsi gli stessi creatori delle forme più evolute della A.I. e dei conseguenti sistemi di comunicazione, da Elon Musk a Mark Zuckerberg, da Sam Altman a Greg Brockman. Torna quindi in campo a pieno titolo anche la filosofia, cioè la valutazione di ciò che l’Uomo vuole essere, piuttosto che l’analisi dell’essere per come è, o meglio per come appare. In tal caso, si può far riferimento al pensiero di Jared Cohen ed Eric Schmidt (2013), i quali a partire dalla loro profonda conoscenza del potenziale di Gen-AI, disegnano alcuni scenari futuri che spaziano dai consumi alla politica, dall’informazione alla cultura, dal conflitto geopolitico ai rapporti tra privacy e sicurezza, avvertendo l’importanza di una costante verifica delle opportunità da cogliere e dei rischi da evitare. E di particolare interesse appare anche la riflessione di John Searle (1985, 2015), il quale ha ribadito la specificità dell’intelligenza umana, capace di una intenzionalità e di una creatività linguistico-concettuale irripetibile da parte dell’A.I. generativa. Searle è l’ideatore del modello esplicativo detto della “stanza cinese”, secondo il quale diversamente dal computer un individuo può elaborare considerazioni utili anche nei confronti di una informazione apparentemente incomprensibile. Su Gen-AI filosofi, sociologi, economisti, informatici hanno elaborato alcune interessanti riflessioni, che tuttavia intrecciano le loro discipline in un complesso rapporto tra “essere” (come ce lo restituiscono le scienze sociali, cioè quelle forme di sapere che analizzano i fenomeni sociali a partire esclusivamente da dati di ricerca empirica teoricamente fondati) e “dover essere” (come lo suggeriscono la filosofia quindi l’etica, magari la logica e finanche la teologia). Il sapere d’altronde non è un sottomarino a compartimenti stagni: al contrario, c’è un continuo scambio tra i dati di ricerca, le conseguenze che se ne traggono, i sistemi di valori che ne tengono conto e i progetti di cambiamento, o di riparazione, che possono innescare.

3. La sociologia e A.I: Costruzione sociale di senso, reti e differenziazione sociale

Di fronte alle domande che si generano quotidianamente intorno a Gen-AI, la sociologia non può giocare un ruolo secondario, ma deve saperne interpretare le valenze, i limiti e le prospettive a livello generale.

In Italia, fra i vari contributi di questo tenore emersi negli ultimi anni possiamo ricordarne tre.

Francesco Morace (2023) descrive una modernità “gassosa”, piuttosto che liquida, nella convinzione che conoscenze, prassi, scelte e possibilità al tempo dell’A.I. si intreccino in un sistema di dipendenze reciproche tridimensionali, tipico della realtà quantistica, offrendo allo stesso tempo nuove opportunità e nuovi rischi. Di qui la necessità di uno strumento versatile e accurato come Gen-AI, in grado di tener conto di tutte le variabili in gioco, della instabilità del loro peso e della loro funzione. Edmondo Grassi (2020), attraverso la ricerca sul campo, matura la convinzione che la realtà costituisca un “sistema” condiviso nel quale l’Uomo non si confronta soltanto con la Natura, ma anche con ciò che egli stesso ha creato – culture, biotecnologie, Intelligenza Artificiale – ed è ormai parte integrante del mondo in cui vive, che oggi è sostanzialmente una “infosfera”. Inoltre, egli sottolinea il bisogno di una “etica”, cioè di un sistema regolativo condiviso in risposta al quesito di come l’Essere Umano intenda vivere la sua esperienza associativa, le sue scelte, le sue pulsioni, tenendosi sempre un passo avanti rispetto ai suoi Alter Ego tecnologici. Massimo Airoidi (2024) coglie invece la profonda influenza che l’Uomo, con la sua cultura, quindi con le sue scelte etiche, ma anche ideologiche e strumentali, esercita sulle Macchine, tal che queste si rivelano esse stesse degli agenti sociali in grado di rinforzare credenze e decisioni: insomma, l’A.I. crea

“abiti digitali” a misura degli orientamenti e delle scelte elaborate dalle collettività, dalle organizzazioni, dai gruppi che ne fanno uso. La ricerca sociologica sembrerebbe in grado di confermare la responsabilità diretta dell’Uomo nelle operazioni, apparentemente autonome, svolte da Gen-AI. Alcuni studi recenti peraltro sembrano alludere alla capacità dell’A.I. di indurre cambiamenti nella condotta individuale e di gruppo, stimolando l’immaginazione umana a concepire sia mondi migliori che futuri distopici. Altre ricerche si orientano sul grado di competenza e di utilizzazione attuale dell’A.I. con riguardo ai singoli individui, ai sistemi mediatici, alle aziende e alle istituzioni. A ben vedere, gli aspetti e i problemi della A.I. Generativa impattano immediatamente con i temi propri della sociologia della conoscenza e in particolare con il processo di costruzione sociale del senso. Senza risalire al pensiero di Max Scheler a di Karl Mannheim su questa basilare condizione, che rende possibile un consorzio civile e un sistema sociale, sarà sufficiente partire dal contributo di Peter L. Berger e Thomas Luckman (1966). Se in termini rousseauiani la società si crea a partire dalla negoziazione e dalla condivisione di leggi ordinarie della convivenza tra individui, è evidente che costoro debbano giungere ad un accordo sull’uso comune di determinati concetti, e principi, a livello ideale e semantico. Ovviamente – e questo lo avevano già suggerito Karl Marx e Frederic Engels – tendono a prevalere quelle definizioni concettuali che vengono proposte e imposte dai detentori del potere; un potere che pur essendo di origine economica, si consolida attraverso un controllo culturale esercitato in termini etici, ideologici e/o religiosi. Tutto ciò è ben noto, ma soprattutto è giunto ad un più vasto pubblico attraverso il pensiero dei sociologi tedeschi dell’Istituto di Francoforte che, in termini marxiani, hanno descritto con toni apocalittici (per dirla con Umberto Eco) i fortissimi cambiamenti culturali avvenuti nel XX secolo in virtù dello smisurato sviluppo della comunicazione,

sia nel campo dell'informazione mediatica che nel campo della propaganda politica e commerciale. Quello che era stato un strumento discreto, ancorché efficace, di influenza sociale da parte delle classi dominanti, attraverso la mitologia eroica delle aristocrazie e la rigida ortodossia imposta dall'autorità religiosa, a partire dal rapido sviluppo della società industriale e delle tecnologie della comunicazione di massa è divenuto uno strumento decisivo per vincere la partita della costruzione di senso del sistema sociale. Che questa impresa sia tentata dal capitale o dal movimentismo rivoluzionario, poco cambia, è una competizione che si vince governando l'informazione, quindi i media di massa; e che soprattutto nella seconda metà del '900, con la crescita delle democrazie liberali e pluraliste, ma anche della globalizzazione e della glocalizzazione, diventa estremamente complessa, articolata, plurale. Le "narrazioni" si moltiplicano; i sempre più numerosi soggetti in grado di esprimersi sui media tentano di riportare la propria visione della realtà, sia che ad agire siano i più potenti network dell'informazione, i governi totalitari o semplicemente l'influencer attivo sui social media con milioni di followers. Ne emerge una ridda di principi, valori, regole, che seppur ordinate gerarchicamente in termini di potenza argomentativa, potere socioculturale, capacità di catturare l'attenzione, influenzare e di persuadere audience estremamente vaste, inevitabilmente coabitano, si sovrappongono e contribuiscono in varia misura alla "definizione sociale" della realtà, mediante complesse procedure negoziali, convenzionali, simboliche. Si vengono dunque a definire moltissime "verità", a vari livelli di attendibilità e di legittimazione. Ci sono le verità ideologiche; quelle religiose; quelle mediatiche; e quelle scientifiche: così la conoscenza umana continua ad essere attraversata da forti contrapposizioni tra "narrazioni" differenti e spesso contrapposte. In particolare, accanto alla scienza sopravvive l'idea, il valore, l'attività filosofica degli individui che non si

accontentano di adeguarsi alle regole del mondo materiale, ma guardano alla dimensione volontaristica insita nella scelta e nella decisione. Questa complessità cognitiva, di natura negoziale, quindi provvisoria e perfettibile, è stata ben rilevata da numerosi sociologi; dalla descrizione di sistemi sociali come sistemi di comunicazione inevitabilmente subordinati all'accordo concettuale, linguistico e semantico, e anche tecnologico, di Niklas Luhmann, a quella già citata della società dell'incertezza di Zygmund Bauman, che forse è soprattutto incertezza cognitiva. L'individuo sceglie da che parte stare; che si tratti di una scelta ideologica, religiosa, filosofica, morale, semplicemente di opportunità e perfino modaiola; ma è chiaro che spesso ha difficoltà a restare nella sua parte di fronte ad un mondo che cambia, propone alternative e nuovi punti di vista (di qui il fenomeno, largamente studiato, di *confirmation bias*). È chiaro come, in questa situazione di incertezza, vissuta solo superficialmente dal cosiddetto uomo della strada, ma ampiamente presente sia nell'azione politica che nelle riflessioni più profonde della scienza, si venga a creare un "bisogno" di verità. Ed è altrettanto chiaro che non a tutti, nella cultura razionalista di ispirazione galileiana della modernità, è sufficiente abbandonarsi ad un Dio. D'altronde lo stesso pensiero cristiano moderno si va orientando verso una concezione postconciliare del divino molto più "socio-antropocentrica", da Karl Rahner ad Hans Küng a Vito Mancuso.

Di fronte ad una incertezza che caratterizza anche la costruzione sociale di senso, la A.I. Generativa è in grado di fornire sentenze – scientifiche, tecniche, sociali, economiche, giuridiche – incontrovertibili, non rifiutabili e coerenti nel creare una visione condivisa della realtà? Evidentemente no. È sufficiente tener conto di due aspetti, sui quali si arrovellano sociologi, filosofi, informatici. Il primo, è di carattere comunicativo e ne abbiamo già accennato: Gen-AI è in grado di creare informazione inesistente,

spacciata per vera e dotata di particolare carisma e di capacità persuasiva. Il secondo aspetto è ben più grave e attiene al tipo di razionalità che ci si attende dalla Macchina. Si potrebbe verificare la sostituzione dei principi etici, nella valutazione e nel governo della società, con un criterio di ottimizzazione fondato esclusivamente su una logica efficientista di tipo “razionale”. Fino a che punto la Macchina potrebbe entrare in conflitto con le decisioni dell’Essere Umano che talvolta si affidano più al sentimento, al valore, all’intuizione che alla ragione?

In questo tipo di logica, basata sulla minimizzazione delle difficoltà e la massimizzazione dei risultati pratici – che è propria di molte branche della tecnologia (nei sistemi di automazione ad esempio) – potrebbe diventare razionalmente opportuno eliminare una qualsiasi minoranza che crei ostacoli e problemi all’efficientamento di un sistema sociale ordinato: che si tratti dei cultori del pensiero critico, degli artisti d’avanguardia, dei migranti sui barchini, dei diversamente abili, dei rivoluzionari, dei sentimentali o dei nostalgici del tempo che fu. Il tutto alla luce di una supposta “oggettività” dei meccanismi di funzionamento dell’Intelligenza Artificiale regolativa della nostra realtà sociale. Non siamo ancora a questo livello di sudditanza verso l’oggettivismo razionalista attribuito a Gen-AI, ma è chiaro che una tale possibilità non è affatto da ignorare. Grassi, come abbiamo visto, richiama la necessità di una “etica” regolativa, di natura pattizia, ma, forse occorre andare anche oltre. Paolo Benanti (2018), consigliere di papa Francesco per la A.I. e tra i membri del Comitato sulla A.I., nominato dal segretario generale delle Nazioni Unite Antonio Guterres, indica la necessità di definire una vera e propria “algoritica”, cioè un modello etico che, a fronte dello sviluppo di algoritmi sempre più raffinati, spesso elaborati dalla stessa Gen-AI, imponga moralmente all’Uomo di assumersi la responsabilità di decidere dove fermarsi, affinché il rispetto della libertà e della

dignità umana prevalga nei confronti della logica meramente efficientistica della Macchina. In realtà, una “oggettività” insensibile alla pietas umana allo stato attuale non può esistere e se esiste siamo noi ad esserne responsabili. Siamo noi a creare la Macchina e siamo noi a definire i suoi limiti di discrezionalità, compresa la possibilità, e il rischio, di generare logiche aliene ai nostri principi etici. I noti “tre principi della robotica” ideati decenni fa da Isaac Asimov, che garantirebbero la subordinazione della Macchina all’integrità dell’Essere Umano, non funzionano se è l’Uomo stesso a liberare la discrezionalità della Macchina quando deve rivolgersi ad un “nemico” umano. La scienza è un modo di conoscere la realtà; probabilmente il più efficace, il più condivisibile perché si fonda su una esperienza logicamente e metodologicamente garantita. Ma sulla base della novella concezione probabilistica della scienza e della conseguente natura negoziale e convenzionale del sapere, in contrapposizione con l’oggettivismo spinto della scienza positivista, noi dobbiamo riconoscere che qualsiasi “verità” dovessimo elaborare a proposito di Intelligenza Artificiale, si tratterebbe pur sempre di una costruzione e di una attribuzione di senso tecnico e scientifico sottoposte a variabili di tipo sociale, culturale, etico e ideologico e perciò stesso modificabili e provvisorie. Un altro campo di analisi della sociologia strettamente legato alla A.I., è costituito dallo studio delle reti sociali, che peraltro svolgono un ruolo fondamentale nei processi di costruzione del senso e nella diffusione relazionale dei messaggi della comunicazione di massa. Il tema delle reti sociali per molto tempo è stato confinato prevalentemente alla dimensione psico-sociale piuttosto che sociologica, nella convinzione che si trattasse soprattutto di un fenomeno di natura interpersonale nel quale l’individuo giocava un ruolo decisivo.

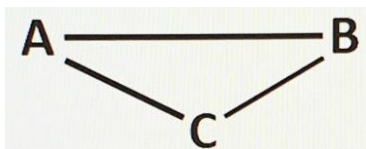
E allora, per comprendere esattamente di cosa si sta parlando, è necessario appellarsi ad alcuni contributi ormai

classici, ma ineludibili perché tuttora fanno scuola. Partiamo quindi dal concetto di inter-azione, cioè di scambio e di relazione interpersonale. Il tema affonda le sue radici sociologiche nel pensiero di Georg Simmel che all'inizio del '900 insiste molto sul concetto di "reciprocità" dei rapporti umani e trova poi uno sviluppo teorico e soprattutto metodologico e di ricerca nella "Sociometria" di Jacob Moreno (1934), che ha come oggetto privilegiato l'"attrazione interpersonale", quella che si stabilisce selettivamente tra alcuni individui, dando vita ad una relazione binaria, sia essa affettiva o di natura funzionale. Questi legami interpersonali spesso si moltiplicano tra più persone, favoriti dalla frequentazione di spazi e di occasioni comuni, e si formano così dei gruppi. Ma la compartecipazione di molti individui (detti "nodi") a più gruppi crea veri e propri network, "reti sociali", che Moreno raffigura mediante grafici detti "sociogrammi". Altri sociologi e antropologi americani e britannici (in particolare la cosiddetta "Scuola di Manchester"), conducendo indagini empiriche sulla vita di relazione in piccole comunità urbane o di lavoro, individuano anch'essi delle "reti sociali" che raggruppano soggetti legati da più intense forme di inter-azione e di collaborazione, sia affettiva, sia funzionale (Mattioli, 2003, 2014). Di qui l'interesse crescente per queste "reti" relazionali che coinvolgono numeri spesso cospicui di soggetti e di qui anche la necessità di introdurre, accanto alla raffigurazione offerta dai sociogrammi, dei sistemi di calcolo matematico per individuare e analizzare più precisamente la consistenza, il funzionamento e le dinamiche di tali reti. L'aspetto che qui ci interessa particolarmente è che queste reti trasportano informazione, cioè sono sì forme di relazione interpersonale e di scambio sociale, ma proprio per questo sono anche reti di comunicazione. Una comunicazione a due vie, in virtù della loro natura interattiva. In seguito, con gli studi dovuti innanzitutto a Stanley Milgram (1967) e a Mark Granovetter (1976), si elaborano

teorie ancor più evolute ed estremamente interessanti per quel che concerne il processo di comunicazione sociale. Ad esempio, Granovetter sottolinea l'importanza dei "legami deboli" che, se non sembrano *bonding* – cioè capaci di rafforzare l'intensità delle relazioni – restano tuttavia *bridging*, perché possono creare ponti con soggetti anche fisicamente lontani fra loro. Da parte sua, Milgram elabora la teoria del "piccolo mondo" e riprende una vecchia e trascurata teoria, detta dei "sei gradi di separazione", dimostrando come sia possibile, utilizzando al meglio le reti, collegare in pochi step soggetti apparentemente estranei. Al declinare del XX secolo due matematici, Duncan J. Watts e Steven Strogatz (1998), dimostrano tecnicamente e matematicamente la giustezza della teoria di Milgram; ma vengono ben presto travolti dallo sviluppo tecnologico dei mezzi di comunicazione. Oggi, con i social media, tu puoi sperare di parlare con chiunque abiti il mondo e abbia accesso a internet, e con molto meno di sei step di intermediazione.

Le "reti sociali" pervadono quindi il vissuto relazionale degli individui e dei gruppi sociali; questa concezione reticolare della società ha convinto definitivamente i sociologi a parlare non più di mera "struttura sociale", ma di "sistema sociale", cioè di una società fondata su una totale interdipendenza delle sue componenti, in continua relazione e dialogo fra di loro. Sia Jeffreys Alexander (1988) che Niklas Luhmann (1984), ad esempio, avvertono che gli elementi costitutivi del sistema sociale interagiscono fra loro in modo dinamico, creando relazioni multiple, instabili, quindi complesse. La preminenza di alcune rispetto ad altre, quindi, non è data una volta per tutte, ma è un fatto storico, culturale, contingente e, per Luhmann, legato alle dinamiche del flusso dell'informazione. Dunque, la società appare come una gigantesca rete di relazioni reciproche e di interdipendenze in costante movimento, non solo tra individui, ma

anche tra soggetti collettivi di varia natura. Proprio perché formandosi assicurano la comunicazione e la collaborazione fra i vari soggetti sociali, le reti contengono e determinano anche le forme negoziali di costruzione sociale di senso; senza di esse, sarebbe difficile non solo concepire i meccanismi di funzionamento di una collettività, ma anche la possibilità di negoziare e di costruire i significati condivisi che sono alla base di una convivenza umana organizzata. Ma c'è un altro punto che va sottolineato. Nelle reti le relazioni acquistano anche un'ulteriore valenza fondamentale: non sono soltanto binarie, o meglio lo sono solo in parte, o apparentemente. Infatti, poiché una relazione si manifesta all'interno di reti più o meno dense e coese, è chiaro che anche la natura, la consistenza e lo sviluppo di tale relazione in qualche misura dipendono dalla rete. Il contesto, e quella che Goffman definisce l'occasione, esercitano una influenza sulla relazione, trasformandola da quella binaria A – B, in quella ternaria così rappresentabile:



Dove A, B, C possono essere individui, gruppi, organizzazioni, culture che in qualche modo si condizionano a vicenda; questo semplice schema ci dice che il rapporto di scambio tra A e B è sempre condizionato da un "esterno" C, che agisce sia sui singoli soggetti che sul loro rapporto, ed è da essi a sua volta condizionato. In effetti, nessuna relazione può essere considerata esclusivamente binaria, se non accordandosi nel sottovalutare o ignorare tutti gli altri fattori che in varia misura intervengono sui due attori e sulle modalità stesse della relazione.

Tutto questo ricorda ancora un altro tema caro alla fisica quantistica: quello, già citato in precedenza, di sovrapposizione o *entanglement*, secondo il quale ogni particella conserva per così dire memoria delle relazioni avute con altre particelle, sì che il suo comportamento relazionale ne è condizionato. In questa prospettiva, Mark Buchanan (2002) si sente autorizzato ad asserire come *tutta* la realtà sia costituita da reti e quindi da componenti in relazione reciproca e dinamica fra loro: in tal caso sia la società umana, sia i sistemi biologici che quelli fisici sono in relazione al loro interno e tra loro, secondo un principio di interrelazione e di interdipendenza che rende difficile continuare a pensare ai meccanismi della realtà, sia quella spirituale intrinseca all'Essere Umano, sia quella materiale del mondo in cui viviamo, come governati da una logica meramente lineare e deterministica. Gli stessi rapporti apparentemente deterministici di causa/effetto sarebbero solo un modo convenzionale e semplificato di concepire pragmaticamente una realtà che è molto più complessa e interrelata per sua stessa natura. Ma c'è ancora dell'altro per il lavoro del sociologo come osservatore privilegiato delle dinamiche sociali. Come si era sottolineato in precedenza, ricordando le dichiarazioni e gli impegni solenni da parte di numerosi Stati Nazionali e da Organismi internazionali come le Nazioni Unite, l'Unione Europea e l'Asean, c'è il rischio fondato di prolungare il *digital divide* che ancora oggi pesa sull'accesso di gran parte dei paesi meno sviluppati alle potenzialità delle nuove tecnologie. Al sociologo spetta il compito di verificare sul campo fino a che punto questi divari siano in grado di condizionare il destino delle popolazioni escluse, quali siano le strade per recuperare lo svantaggio, e quali effetti si possano verificare nel caso di una "democratizzazione" dell'Intelligenza Artificiale, sia che su di essa resti intatto il controllo diretto da parte dell'Essere Umano, sia che lo sviluppo delle tecnoscienze produca prima o poi una

“intelligenza” in grado di prendere decisioni almeno apparentemente autonome. Tuttavia, c’è un’altra forma di differenziazione sociale che interessa direttamente la sociologia, con riguardo alla capacità di maneggiare l’A.I. nei paesi altamente sviluppati: si tratta della differenziazione generazionale. Sebbene siano stati i *babyboomers* ad avviare l’Intelligenza Artificiale e siano state le Generazioni X e Y a renderla una risorsa comune al girare del nuovo Millennio, sono le generazioni nate nel XXI secolo (definite non a caso *A.I. Generation*), a vivere la presenza della A.I. come parte integrante sia della quotidianità attuale che di quella futura. Il loro linguaggio, il loro tempo libero e le loro forme di associazionismo, i loro valori – quali che essi siano, da quelli consumistici a quelli ideologici, da quelli religiosi a quelli edonistici – viaggiano lungo le reti e secondo le logiche dei social media. Loro sono i veri abitanti dell’“infosfera”, in grado di spezzare i limiti della circolazione della comunicazione, mutandone la forma, la consistenza, la qualità, le procedure, i supporti. Sono loro a considerare normali le potenzialità del metaverso. Tutto questo a dimostrazione, peraltro, che il possesso di competenze e di mezzi intellettuali non è meno determinante del possesso dei mezzi di produzione e che oggi il cambiamento culturale, cioè quello che attiene al costume, alle credenze, ai valori può essere altrettanto determinante. Anche perché, oggi, il cambiamento sembra correre come non mai e ne siamo noi i responsabili. Sta acquistando sempre più valore, infatti, la vecchia “Legge di Moore”, quella che già qualche decennio fa sottolineava la rapidità dei cambiamenti in atto nella tecnologia e che oggi descrive la tendenza di tali ritmi a presentarsi sempre più serrati e in grado di innovare e perfezionare quasi totalmente i sistemi tecnologici disponibili nel giro di meno di due anni. Possiamo dunque trarre qualche conclusione. In primis, che il mondo in cui viviamo è solo quello che a noi appare per come lo sperimentiamo, dalla consistenza degli oggetti ai loro colori,

dalla misurazione del tempo alle manifestazioni che consideriamo vitali, dalle regole che ne traiamo alla valenza creativa dell'operato umano. Ed è sulla base di questi nostri "punti di vista" intellettuali e sensoriali che noi lo definiamo e lo interpretiamo. E questo vale anche per la A.I. Anzi, forse oggi soprattutto per la A.I. Perché anche le sue attribuzioni e le sue reali funzioni dipenderebbero da numerose variabili esterne che agiscono di concerto fra loro e con la Macchina, che vanno dalla banale disponibilità dei supporti materiali necessari (metalli rari, energia) alla fondamentale volontà dei programmatori umani e alle loro intenzioni. Su questo, la sociologia offre agli studiosi che si occupano delle possibilità, delle prospettive e dei rischi dell'A.I. alcuni punti di riferimento teorici ed empirici fondamentali: a) la determinazione sociale della conoscenza intesa come attribuzione e costruzione condivisa di senso; b) i caratteri negoziali della conoscenza e segnatamente della scienza; c) la funzione costitutiva delle reti sociali nel processo di comunicazione e in specie della creazione e della diffusione dell'informazione. Il tutto perché la A.I. genera rilevanti effetti proprio in queste dimensioni della vita umana. Rispetto al punto a) occorre definire come si comporta l'Intelligenza Artificiale Generativa in una società umana che attribuisce un senso negoziato alla realtà: adattandosi ad esso o producendone uno nuovo? In tal caso, la società lo accetta inchinandosi alla razionalità tecnologica o rivendica una originalità propria del pensare umano? E a quali livelli di potere si gioca questa partita?

Per il punto b) si va delineando una competizione tra due diversi "partiti" della scienza: da un lato coloro che non tollerano alcun controllo sullo sviluppo del sapere, sia ricordando i tempi bui del medioevo, sia temendo l'invasione o magari lo sfruttamento strumentale da parte della politica; all'opposto si pongono gli scienziati che temono il sorgere di una tecnocrazia che poco o nulla avrebbe da invidiare ad una teocrazia o addirittura l'insorgere di

un Crono capace di divorare i propri figli. Infine, per il punto c): è un dato di fatto che oggi, proprio grazie ad internet e alla A.I. la rete delle relazioni sociali e delle comunicazioni ha preso una svolta mai vissuta in precedenza: l'estensione delle relazioni è ormai a livello globale, e senza neppure tanti intermediari, e trasforma i rapporti sociali attenuandone forse l'intensità, ma aumentandone la rilevanza collettiva; intanto nuovi soggetti sociali riescono a far sentire la propria presenza, intervenendo nel processo di costruzione di senso proprio a partire dalla loro attivismo in rete; infine, la costruzione del vero può essere modificata dalla costruzione condivisa del verosimile e del desiderabile. Così, forse, è proprio la versatilità dell'Intelligenza Artificiale – a cui oggi ci si affida per la sostenibilità ambientale, per la giustizia sociale, per l'efficienza produttiva e quella organizzativa, per le forme più trasversali della comunicazione, per una più efficace tutela della salute, per un sistema educativo e formativo migliore, per un arricchimento della conoscenza e della scienza – a generare, con la sua molteplicità di funzioni e di opportunità, anche una ulteriore e ancor meno prevedibile complessità. Sta all'Uomo trovare la risposta adatta ai problemi emergenti affinché il progresso non si trasformi in una minacciosa involuzione; e non è la prima volta che accade, nella recente modernità. È accaduto infatti con l'energia nucleare, utile sì in moltissimi casi, ma che ha costretto l'Umanità ad autoregolarsi affinché nella sua espressione civile ma soprattutto militare non divenisse un pericolo planetario che non avrebbe fatto distinzione tra vincitori e vinti. Sta accadendo con la minaccia climatica e ambientale, che costringe a trovare soluzioni alternative all'uso indiscriminato delle risorse energetiche non rinnovabili e fortemente inquinanti per l'atmosfera e le acque planetarie. E potrebbe accadere con una Gen-Ai male impostata e quindi incontrollata che imponesse una "verità" computazionale distinta

da un'etica, certamente perfettibile, negoziabile, ma comunque elaborata dalla "singolarità umana".

Bibliografia

Airoidi M. (2024). *Machine habitus. Sociologia degli algoritmi*. Roma: Luiss University Press.

Alexander J. (1988). *Action and His Environments. Toward a New Synthesis*. New York: Columbia University Press.

Bauman Z. (2000). *Liquid Modernity*. Cambridge: Polity Press. (tr.it. *Modernità liquida*, Roma-Bari, Laterza, 2003).

Benanti P. (2018), *Oracoli. Tra algoretica e algocrazia*. Roma: Luca Sossella Editore.

Berger P.L., Luckmann T., (1966) *The Social Construction of Reality*. New York: Doubleday (tr. it. *La realtà come costruzione sociale*, Bologna, Il Mulino, 1969).

Buchanan M. (2002), *Nexus. Small World and the Groundbreaking Science of Networks*, New York, Norton. (tr. it. *Nexus. La rivoluzionaria teoria delle reti*, Milano, Mondadori, 2003).

Cohen J., Schmidt E. (2013). *The New Digital Age. Reshaping the Future of People, Nations and Business*. Murray (tr. it. *La nuova era digitale. La sfida del futuro per cittadini, imprese, nazioni*, Milano, Rizzoli, 2013).

Granovetter M. (1976). *Getting a Job*. Cambridge MA: Harvard University Press.

Grassi E. (2020). *Etica e Intelligenza Artificiale. Questioni aperte*. Roma: Aracne.

Kuhn T. (1962). *The Structure of Scientific Revolution*. Chicago: Chicago University Press (tr.it. *La struttura delle rivoluzioni scientifiche*, Torino Einaudi, 1978).

Luhmann N. (1984). *Soziale Systeme*. Frankfurt am Main: Suhrkamp (tr. it. *Sistemi sociali. Fondamenti di una teoria generale*, Bologna, Il Mulino, 1990).

Mattioli F. (2003). *Sociometria*, III ediz. Roma: La Goliardica.

Mattioli F. (2017), La network analysis tra teoria e ricerca. In Mattioli F., Anzera G., Toschi L., *Teoria e ricerca nell'analisi delle reti sociali*. Roma: Aracne. 11-54.

Merton R.K. (1957). *Theory and Social Structure*. Glencoe: Free Press. (tr. it. *Teoria e sociale*, Bologna, Il Mulino, 1992).

Milgram S. (1967). The Small World Problem. *Psychology Today*, 1(1), 61-67.

Morace F. (2023). *Modernità gassosa*. Milano: Egea.

Moreno J. (1934). *Who Shall Survive?* New York: Beacon Press. (tr.it. *Principi di sociometria, di psicoterapia di gruppo e sociodramma*, Milano, Etas Kompass).

Popper K. (1935). *Logik der Forschung*. Wien: Springer (tr. ingl. *The Logic of Scientific Discovery*, London, Hutchinson, 1959; tr. it. *Logica della scoperta scientifica*, Torino, Einaudi, 1970).

Roncaglia G. (2023). *L'architetto e l'oracolo. Forme digitali del sapere da Wikipedia a ChatGTP*. Roma-Bari: Laterza.

Searle JR. (1980). Minds, Brains and Programs. *Behavioral and Brain Sciences*, 3(3), 417-424. doi:10.1017/S0140525X00005756.

Taleb N.N. (2005). *The Black Swan. The Impact of Higly Improbable*. New York: Random House (tr. it. *Il cigno nero*, Milano: Il Saggiatore, 2008).

Watts D.J., Strogatz S.H. (1998). *Collective Dynamics of Small-World Networks*, *Nature*, 393, 440-442.

2. INTELLIGENZA ARTIFICIALE E LITERACY. PROMUOVERE L'APPROCCIO SOCIOLOGICO UMANO- CENTRICO PER SUPERARE I PREGIUDIZI E FAVORIRE L'INCLUSIONE SOCIALE

di Danilo Boriati*, Mariangela D'Ambrosio**

Abstract: Il saggio, partendo da un'analisi sociologica sul ruolo dell'intelligenza artificiale nella società contemporanea, focalizza la propria attenzione sull'alfabetizzazione digitale e sulla funzione della *digital literacy* quale capacità di utilizzare le tecnologie digitali in modo efficace e critico, in un'attenzione che promuova un uso consapevole delle tecnologie e che rimetta al centro l'uomo come agente dotato di coscienza, valutazione e interpretazione.

Parole chiave: IA, pregiudizi, approccio umano-centrico, alfabetizzazione digitale, inclusione sociale.

Abstract: Starting from a sociological analysis of the role of artificial intelligence in contemporary society, the essay focuses on digital literacy as the ability to use digital technologies efficiently and critically, in a way

* Facoltà di Scienze della Comunicazione, Università Telematica Internazionale Uninettuno, orcid.org/0009-0001-4496-4325, danilo.boriati@uninettunouniversity.net.

** Dipartimento di Economia, Università degli Studi del Molise, orcid.org/0000-0002-8147-5299, mariangela.dambrosio@unimol.it.

Seppure il lavoro sia il frutto di una riflessione congiunta, in termini formali l'introduzione, i paragrafi 3, 3.1 e le conclusioni vanno attribuiti ad entrambi gli autori, mentre i paragrafi 1 e 2 a Danilo Boriati e i paragrafi 4 e 5 a Mariangela D'Ambrosio.

Accettato Dicembre 2024 - Pubblicato Aprile 2025.

that promotes a conscious use of technologies and puts humans at the centre as agents endowed with consciousness, evaluation, and interpretation.

Keywords: AI, biases, human-centric approach, Digital Literacy, social inclusion.

Introduction

L'Intelligenza Artificiale (d'ora in poi IA) sta trasformando rapidamente la società in modo profondo e pervasivo. Si tratta, infatti, di un processo di cui gli individui-utenti (*prosumer*, dunque *consumer* e *producer* al tempo stesso) hanno ancora poca reale contezza. Un cambiamento che si manifesta in diversi settori della vita, sia individuale sia aggregata, che se, da un lato, può comportare molti benefici, dall'altro pone in campo importanti questioni di carattere etico. Da un punto di vista propriamente sociologico, l'IA rappresenta un sistema socialmente costituito (Schwartz, 1989), che costruisce senso (Berger e Luckmann, 1966) quando applicata a un contesto specifico dove essa stessa può assumere ruoli sociali (Beer, 2017), creare vere e proprie relazioni sociali nelle quali, a sua volta, interagisce, e dove riproduce *bias* (Bechmann e Bowker, 2019; Broussard, 2018; Obermeyer *et al.*, 2019; Parikh *et al.*, 2019; Franzoni, 2023), discriminazioni (Bloomfield, 1987; Fleck, 1987; Forsythe, 1993a, 1993b) e stereotipi (Varsha, 2023). Gli “stili di pensiero”, come insieme di credenze e convinzioni, gravano in tal senso sulla percezione della realtà sociale, influenzando opinioni e comportamenti nella vita quotidiana (Leavy, 2018) perché generalizzabili e, spesso, caratterizzati da quei pregiudizi le cui conseguenze possono essere molto negative. In questo quadro, il presente contributo vuole sottolineare il compito decisivo e strategico della *Digital Literacy*

quale capacità di utilizzare le tecnologie digitali in modo efficace e critico (D'Ambrosio e Boriati, 2023), valutando (e creando al contempo) contenuti digitali in un'attenzione che "riporta" al centro l'uomo come agente dotato di coscienza, valutazione e interpretazione (in termini di *Human Centered AI*). Attivare, dunque, il pensiero critico, allenarlo e maneggiare le possibili distorsioni connesse all'interiorizzazione e alla riproduzione di *habitus* (Bourdieu, 1975) può consentire un uso consapevole del sistema di credenze e di aspettative, nonché l'instaurazione di relazioni sociali adatte a navigare con successo nell'era digitale, nell'ottica di creare prosperità collettiva. Un esercizio di accompagnamento all'IA – un'*Artificial Intelligence Literacy* (Cuomo *et al.*, 2022) che diventa necessario, dunque, per promuovere la partecipazione civica, l'inclusione sociale e l'equo sviluppo della società e, dunque, anche dei sistemi di Welfare – che deve, soprattutto nel momento storico e politico attuale, essere in grado di superare le già presenti ma anche le inedite ed emergenti problematiche connesse all'uso dell'IA.

1. Il ruolo dell'IA nella società post-moderna. Brevi riflessioni preliminari

Da diversi anni è in atto una rivoluzione tecnologica senza precedenti, che sta letteralmente cambiando il mondo. Il rapido sviluppo delle tecnologie in generale – e dell'IA in particolare – sta plasmando, con sempre maggiore incisività, la civiltà contemporanea. Tale sviluppo ha indotto alcuni autori a parlare della società post-moderna in termini di *platform society* (van Dijck *et al.*, 2019), una società in cui vi è una «inestricabile relazione tra le piattaforme on-line e le strutture sociali» (ivi:24) e dove forti sono i condizionamenti delle piattaforme digitali sui molteplici aspetti

della vita quotidiana, dell'economia e della cultura. Si tratta di un *topoi* che concepisce l'innovazione massiva delle tecnologie digitali non solo come un processo meramente tecnologico, quanto piuttosto come un fondamentale mutamento socioculturale. Questa pressante rivoluzione digitale, infatti, influisce sulle scienze sociali in vari modi. In prima istanza, i processi sociali subiscono cambiamenti e adattamenti fondamentali che richiedono un'ampia elaborazione scientifica. In secondo luogo, l'applicazione sempre più diffusa delle tecnologie digitali genera un'enorme massa di dati (georeferenziati, di tracciamento, di testo dei social media) (Golder e Macy, 2014). In terzo luogo, la crescente potenza di calcolo e la maturazione degli ambienti software hanno promosso lo sviluppo di soluzioni algoritmiche per problemi statistici complessi. Questo ha aperto la strada al nascente campo delle scienze sociali computazionali (Edelmann *et al.*, 2020) e all'intersezione tra scienze sociali e statistica, informatica e matematica (Leitgöb *et al.*, 2023). In questo quadro di forte cambiamento della società contemporanea, l'IA sta giocando un ruolo sempre più centrale nel determinare la formazione di un contesto nel quale il potere degli algoritmi è diventato preminente, tanto da produrre molteplici discussioni all'interno della società, dei media e delle stesse scienze sociali. L'IA e i sistemi di apprendimento automatico, infatti, vengono sovente descritti come capaci di prendere decisioni più intelligenti, più veloci, migliori e presumibilmente più neutrali: a tal proposito, Yeh *et al.* (2021) sostengono che l'IA sia pronta a rivoluzionare la vita degli attori sociali. In continuità, l'Unione Europea, insieme ai governi di tutto il mondo, vede l'IA come la più grande promessa del ventunesimo secolo, in grado di apportare contributi positivi a molteplici aspetti della vita umana, dal miglioramento dell'assistenza sanitaria alla mitigazione dei cambiamenti climatici (Commissione Europea, 2020). Allo stesso tempo, però, viene sottolineato nel dibattito scientifico come i

sistemi di IA comportino altresì una serie di rischi, come pregiudizi e discriminazioni di genere o anche connessi all'età, nonché la violazione del diritto alla privacy e la promozione di limitazioni alla libertà personale (Rahwan, 2018; Tufekci, 2014). Gli studiosi di IA etica, a tal proposito, esprimono preoccupazione sul modo in cui queste tecnologie mostrano pregiudizi nascosti che portano all'esclusione e alla discriminazione dei membri di alcuni gruppi sociali, già di per sé marginalizzati (Eubanks, 2018; Mittelstadt *et al.*, 2016; Lloyd, 2018), rappresentando in tal modo persino una minaccia ai diritti umani fondamentali e alla giustizia sociale (Cruz, 2020). È indubbio, tuttavia, come il panorama delle riflessioni epistemologiche sul tema dell'IA e sul suo impatto all'interno della società post-moderna si stia ponendo non tanto dal punto di vista del processo di accesso alle informazioni quanto piuttosto su quello, decisamente più complesso, di costruzione e mantenimento delle relazioni sociali. In questo quadro, dunque, i tentativi di esaminare il rapporto tra le tecnologie digitali come l'IA e la società hanno portato a una nuova descrizione dell'ordine sociale, nonché delle possibilità, dei rischi e delle sfide attuali e future poste da questi strumenti.

2. Per una Sociologia dell'IA: tra diffusione, ricezione e utilizzo delle nuove tecnologie

Parlare di sociologia dell'IA vuol dire, ad avviso di chi scrive, offrire uno strumento fondamentale per analizzare criticamente il ruolo dell'IA nella trasformazione delle strutture sociali. Così come avvenuto per le altre tecnologie (internet, social media, etc.), anche per l'IA diviene possibile – adottando le lenti proprie della scienza sociologica – analizzare tale tecnologia come fattore di cambiamento sociale (McLuhan, 1967; Castells, 1996) da un lato e

come fattore influente su cultura e società e viceversa dall'altro (Hughes, 1987; Winner, 2020). Infatti, l'IA non è solamente una tecnologia neutra, ma può essere considerata come una forza che modella e che viene a sua volta modellata dalle dinamiche sociali, economiche e politiche; esplorare queste interazioni da un punto di vista sociologico può aiutare a comprendere come la società possa affrontare in modo equo ed etico le sfide che l'IA pone per il prossimo futuro. All'interno di un'analisi sui prodotti scientifici inerenti l'IA, Liu (2021) individua tre direttrici della riflessione sociologica, sottolineando come ciascuna delle prospettive si concentri sull'IA come campo di ricerca scientifica, sull'IA come applicazione sociale e sull'IA come causa di effetti economici, politici, sociali e culturali. Nel presente lavoro, l'attenzione è posta proprio su quest'ultima direttrice, ovvero sulla IA culturale (Liu, 2021:8), poiché essa – considerando l'IA come fenomeno prettamente sociale – permette di porre luce sulla discriminazione che i nuovi artefatti tecnologici possono porre in essere. L'IA, proprio in quanto artefatto tecnologico, “segue” le stesse dinamiche sociali e culturali di diffusione e ricezione della tecnologia *tout court*. Due sono i modelli sociologici che permettono di descrivere i fenomeni di incorporazione delle tecnologie da parte degli attori sociali: diffusionismo da una parte e addomesticamento dall'altra. La teoria della diffusione delle innovazioni tecnologiche (Rogers, 2003) presuppone che alla base dell'utilizzo di una innovazione vi sia un processo comunicativo e relazionale che permette a quest'ultima di entrare a far parte della vita quotidiana dei suoi utilizzatori. La teoria dell'addomesticamento (Silverstone e Hirsch, 1992), invece, si interroga su come la tecnologia venga a far parte, in maniera effettiva, della vita quotidiana di individui e famiglie e di come questa si collochi nel più ampio sistema mediale casalingo. In questo quadro epistemologico, pertanto, diviene possibile analizzare l'IA non solamente nei termini di modificazione delle

strutture sociali vigenti, ma anche su come le persone percepiscono sé stessi e le loro relazioni, tanto con altri individui quanto con la tecnologia stessa. Le interazioni quotidiane con dispositivi IA, come assistenti virtuali, chatbot e sistemi automatizzati, stanno difatti creando nuove modalità di interazione sociale. A tal proposito, è importante sottolineare come l'uso quotidiano dell'IA potrebbe comportare, per l'attore sociale contemporaneo, il rischio di cedere «il passo al dominio della tecnologia, facendosi usare dalla macchina. Si tratta di un rischio di cui si parlerà a lungo, finché l'Intelligenza artificiale non avrà raggiunto una sua stabilità e un grado di conoscenza e consapevolezza tale da considerarla, più che una minaccia, un *modus vivendi* che può migliorare la centralità della persona, piuttosto che annullarne la presenza» (Centorrino e Romeo, 2023:107). Nel più ampio contesto dell'era digitale, l'IA sta dunque diventando una forza trainante che modella la società in modi sempre più profondi. Per affrontare le sfide e le opportunità che l'IA presenta e per evitare contestualmente che la persona perda la propria centralità, è essenziale sviluppare una competenza chiave: la *Digital Literacy*. Questo concetto va oltre la semplice capacità di utilizzare la tecnologia, in quanto implica una comprensione critica di come l'IA funziona, delle sue implicazioni etiche e sociali, e della capacità di partecipare attivamente al dibattito pubblico sull'uso e sullo sviluppo di queste tecnologie.

3. L'importanza della Digital Literacy nell'IA: strumenti per comprendere la società e partecipare nel mondo

La *Digital Literacy* nell'IA può essere definita come la competenza che permette agli individui di comprendere, analizzare e utilizzare in modo critico le tecnologie basate sull'intelligenza

artificiale. Secondo Long e Magerko (2020), questa competenza non si limita alla conoscenza tecnica degli algoritmi o dei modelli di apprendimento automatico, ma include anche e soprattutto la comprensione delle implicazioni etiche, sociali ed economiche dell'IA. In particolare, la *Digital Literacy* può essere definita come l'insieme di competenze, conoscenze e abilità necessarie per utilizzare le tecnologie digitali in modo efficace, sicuro e critico. Competenza che, come accennato, non si limita alla capacità di utilizzare strumenti digitali, ma include anche la capacità di comprendere, valutare e creare contenuti digitali, gestire la propria identità *online* e partecipare attivamente nella società digitale. La *Digital Literacy* implica, inoltre, la consapevolezza delle implicazioni etiche e sociali delle tecnologie digitali e la capacità di navigare tra le informazioni in modo critico, riconoscendo la credibilità delle fonti e la validità delle informazioni (Gilster, 1997; Eshet-Alkalai, 2004). Sulla questione, risulta importante quanto asserito da Prensky (2001), il quale, alle categorie dei *digital natives* e *digital immigrants* – da lui stesso create –, preferisce quelle di *digital skilness*, *digital wisdom* e *digital stupidity* (2009, 2010). Nello specifico, la *digital skilness* si riferisce alla capacità di padroneggiare efficacemente le tecnologie digitali: essa include le competenze tecniche necessarie per utilizzare dispositivi mediiali e piattaforme *online* in modo efficiente e produttivo. Essa non è, quindi, soltanto l'avere accesso alla tecnologia, ma anche di saperla usare in modo competente. Ed ecco, dunque, perché si parla anche di *digital wisdom*, o saggezza digitale: l'uso delle tecnologie digitali deve avvenire con giudizio critico e una comprensione etica e critica. Chi è dotato di *saggezza* non si limita a navigare nel mondo digitale, ma lo fa con consapevolezza, considerando le conseguenze, i rischi e i pericoli in termini di *privacy*, corretto utilizzo delle informazioni, gestione dell'identità digitale e capacità di discernere tra informazioni affidabili e

disinformazione. In coerenza, si può verificare la *digital stupidity*, o stupidità digitale, quando si adoperano in maniera irresponsabile e imprudente le tecnologie digitali e, oggi, anche l'IA. Essa si riferisce ad una mancanza di giudizio nel modo in cui la tecnologia viene utilizzata, spesso con conseguenze negative per sé e per gli altri. Sulla stessa scia, come ben sottolineato da Noble (2018), la mancanza di consapevolezza critica riguardo all'IA può portare a una perpetuazione delle disuguaglianze sociali; Floridi (2019), invece, evidenzia come una partecipazione informata sia cruciale per garantire che lo sviluppo dell'IA sia allineato con i valori e i principi umani. In tal senso, va ricordato che si tratta del più ampio esercizio della cittadinanza digitale: in un mondo dove i sistemi informatici intelligenti hanno un impatto sempre più pervasivo, l'*Artificial Intelligence Literacy* emerge come una competenza essenziale, per comprendere la società e partecipare attivamente nel mondo, che «deve articolarsi su dimensioni conoscitive ed operative, ma anche critiche ed etiche. Inoltre, per formare cittadini responsabili in un mondo ad alta densità tecnologica, appare fondamentale progettare e sviluppare progetti di alfabetizzazione all'IA [...], favorendo la partecipazione del maggior numero di persone» (Cuomo *et al.*, 2022:168) possibili. Investire, quindi, nell'educazione alla *Digital Literacy* – nei sistemi di IA – è, dunque, fondamentale per costruire una società più equa, consapevole e soprattutto democratica.

3.1 L'approccio umano-centrico e la competenza digitale: per una promozione globale dei valori, dei bisogni e dei diritti umani nell'esercizio del pensiero critico

Nell'era della trasformazione digitale, l'approccio umano-centrico, ponendo l'essere umano al centro dei processi di innovazione e sviluppo tecnologico, richiede una comprensione

profonda delle tecnologie digitali e delle loro implicazioni, anche e soprattutto nell'esercizio della "competenza digitale" quale uno fra gli strumenti necessari per affrontare le sfide della società contemporanea. Tale aspetto diventa necessario soprattutto per Paesi come l'Italia, che si posiziona agli ultimi posti in UE per competenze digitali a cui si aggiunge il divario di genere e geografico (Istat, 2023): «le competenze digitali [...] sono caratterizzate da un forte divario di genere a favore degli uomini, che, nel nostro Paese è di 5,1 punti percentuali» (*ibidem*).

L'approccio umano-centrico applicato all'IA – che si pone come ulteriore sfida nel sistema più ampio delle competenze digitali, ad oggi necessario – è una filosofia di progettazione e implementazione che pone l'essere umano al centro del processo tecnologico. Si tratta di un approccio che enfatizza l'importanza di sviluppare sistemi di IA che rispettino, promuovano e migliorino i valori, i bisogni e i diritti degli esseri umani. Secondo Norman e Draper (1986), tale orientamento si concentra sulla comprensione profonda degli utenti, dei loro contesti e delle loro esigenze, cercando di creare soluzioni che siano intuitive, accessibili e che rispettino i valori umani, invece di concentrarsi esclusivamente sulle capacità tecniche e sugli obiettivi economici; un approccio umano-centrico, in definitiva, considera le implicazioni sociali, etiche e culturali della tecnologia. Per integrare efficacemente la competenza digitale nell'approccio umano-centrico, infatti, è necessario un equilibrio tra competenze tecniche e competenze umanistiche. Da un lato, sviluppatori e utenti devono possedere una solida comprensione delle tecnologie emergenti, come l'*Artificial Intelligence* e l'*Internet of Things*. Dall'altro lato, essi devono essere consapevoli delle implicazioni etiche del loro utilizzo, come il rispetto della *privacy* e della sicurezza, nonché l'uso dell'etica per l'inclusione sociale di tutti. In questo contesto, e proprio rispetto al tema dell'etica, vale la pena aprire una breve

parentesi sulla proposta della convenzione internazionale giuridicamente vincolante riguardante l'IA che ha rappresentato, recentemente, un passo cruciale per regolamentare lo sviluppo e l'uso di queste tecnologie, assicurando che esse siano impiegate in modo responsabile e al servizio del bene comune. Si tratta della Convenzione Quadro del Consiglio d'Europa sull'intelligenza artificiale, i diritti umani, la democrazia e lo Stato di diritto, che definisce l'uso dell'IA a livello mondiale «con l'obiettivo di promuovere l'innovazione e l'adozione di un'IA affidabile, proteggendo al tempo stesso la salute, la sicurezza e i diritti fondamentali, compresi la democrazia, lo Stato di diritto e l'ambiente» (CoE, 2003:3). L'obiettivo centrale è, d'altra parte, «garantire che le attività che rientrano nel ciclo di vita dei sistemi di IA siano pienamente coerenti con i diritti umani, la democrazia e lo Stato di diritto» (Ivi:4). Le attuali normative nazionali e regionali non sono sufficienti per affrontare le sfide globali poste dall'IA, rendendo indispensabile un accordo internazionale che stabilisca standard comuni (Cath *et al.*, 2018). Senza un quadro normativo internazionale, invero, i rischi associati all'IA, come la discriminazione algoritmica, la violazione della *privacy* e l'uso improprio e/o in contesti bellici, potrebbero esacerbarsi (Floridi, 2018). Certo, l'adozione di una convenzione internazionale sull'IA non è priva di sfide. Una delle principali difficoltà risiede nelle differenze culturali, politiche ed economiche tra i vari Paesi, che potrebbero portare a disaccordi sui principi e le norme da adottare. Inoltre, la natura in continua evoluzione della tecnologia rende difficile prevedere e normare tutte le possibili applicazioni dell'IA. Come evidenziato da Walsh *et al.* (2020), una regolamentazione troppo rigida potrebbe soffocare l'innovazione, mentre una troppo permissiva potrebbe lasciare spazio ad abusi. Una dimensione, quindi, complessa che richiama – come anticipato – da parte degli utenti l'essere competenti digitalmente, ossia essere capaci di

usare con dimestichezza e spirito critico le tecnologie della società dell'informazione (Commissione Europea, 2018): ciò non si limita all'uso di strumenti digitali, ma include la capacità di comprenderne gli impatti socio-economici e culturali. Essere competenti digitali, in maniera specifica, implica l'uso sicuro, critico e responsabile delle tecnologie digitali e il loro impiego nell'apprendimento, nel lavoro e nella partecipazione alla società. Elementi che si collocano nella promozione e nel rispetto globale dei valori, dei bisogni e dei diritti umani. L'approccio umano-centrico, quindi, insieme alla competenza digitale, rappresenta una delle possibilità per migliorare la qualità della vita e rispettare la dignità e i diritti individuali, soprattutto (ma non esclusivamente) tramite lo sviluppo del pensiero critico, in quanto la capacità di pensare criticamente è diventata essenziale per navigare tra le complessità dell'IA e delle tecnologie digitali. Il pensiero critico, definito come l'abilità di analizzare, valutare e formare giudizi informati su questioni complesse, gioca un ruolo fondamentale nell'interazione con le tecnologie avanzate, così come previsto proprio dal DigComp 2.2 (*Digital Competence Framework for Citizens*). Il pensiero critico non si limita, in altre parole, alla mera acquisizione di informazioni ma implica la capacità di mettere in discussione, interpretare e riflettere su ciò che si apprende, in un processo di autovalutazione delle proprie convinzioni e azioni, che richiede la sospensione del giudizio e l'apertura al cambiamento (Facione, 2011). Una competenza essenziale in un mondo dove l'IA e le tecnologie digitali influenzano ogni aspetto della vita quotidiana: dalla formazione delle opinioni all'automazione del lavoro, dalla formazione del pensiero al prendere decisioni individuali (dalle ricadute collettive). L'intelligenza artificiale, va sottolineato, con la sua capacità di elaborare grandi quantità di dati e prendere decisioni complesse, pone una sfida mai vista prima rispetto al pensiero critico. Gli algoritmi di IA, sebbene

potenti, non sono esenti da pregiudizi e limitazioni e, di conseguenza, ecco perché risulta stringente che gli utenti sviluppino un pensiero interpretativo e globale per comprendere e valutare le decisioni prese dalle “macchine”, anche in termini di pregiudizi impliciti, ambiguità e iniquità (Binns, 2018). In un mondo sempre più dominato dalla tecnologia la competenza digitale rappresenta, pertanto, una componente essenziale per l’approccio umano-centrico (e viceversa). La fusione di competenze tecniche e sensibilità etiche è cruciale per sviluppare tecnologie che siano non solo efficaci, ma anche rispettose dei valori umani. Solo attraverso una comprensione approfondita delle tecnologie e delle loro implicazioni potremmo assicurare che l’innovazione digitale rimanga al servizio dell’umanità, promuovendo un progresso che sia inclusivo e sostenibile.

4. Superare i pregiudizi. Strategie e applicazioni positive dell’IA nell’“onlife”

Nell’“onlife” l’IA gioca un ruolo sempre più centrale, non mancando di preoccupazioni legate, fra le altre, ai pregiudizi algoritmici che possono perpetuare o amplificare le disuguaglianze sociali già presenti nel contesto societario. Va precisato che, per “onlife”, s’intende una sintesi della dimensione reale con quella virtuale, dove i confini sono sempre più labili in una compressione spazio-temporale sempre più marcata (Floridi, 2015). La metafora delle mangrovie, in questo senso, suggerisce che proprio come questi arbusti vivono immersi in acqua salmastra (a metà fra acqua dolce e acqua salata), così noi tutti esistiamo e, dunque, agiamo, in un mondo digitale e pragmatico allo stesso tempo. Nella dimensione ibrida qui descritta, (quasi) tutte le azioni quotidiane sono interrelate tra di loro in un’estensione contemporaneamente

reale e virtuale che, in termini di categorizzazione tradizionale, sarebbe bene, ormai, evitare del tutto (Hildebrandt, 2015). Nel quadro così delineato, allora, i pregiudizi (di qualunque natura e tipologia essi siano) non solo persistono ma possono anche essere amplificati dalle tecnologie digitali, in particolare dall'IA, dagli algoritmi e dal *machine learning*, in un rimando reciproco che riproduce socialmente elementi stigmatizzanti con il rischio di escludere precise categorie di persone (Arrow, 1973). Si pensi, ad esempio, ai sistemi attuali di recruiting che spesso possono essere "tarati" sulla ricerca di una certa tipologia di personale (Köchling e Wehner, 2020; Hofeditz *et al.*, 2022), oppure si pensi alla più diffusa discriminazione basata sul genere, l'età, l'orientamento religioso e sessuale e sulle abilità/capacità (Lee, 2018); o ancora, alla trasmissione di determinati valori e norme sociali (Baker e Hawn, 2022). I pregiudizi sono spesso il risultato di abitudini mentali statiche e di un'eccessiva generalizzazione. E, nel contesto "onlife", queste dinamiche sono esacerbate dalla natura pervasiva delle tecnologie digitali che "filtrano" e organizzano le informazioni in modi che possono rinforzare le convinzioni preesistenti, riproducendo bias (Kahneman *et al.*, 1982). Nel contesto sociale, l'insieme delle inclinazioni, delle attitudini nonché degli schemi di pensiero, sono il frutto di condizionamenti sociali che incidono sui pensieri e sulle scelte degli individui, in una struttura di potere spesso latente (Bourdieu, 1998). L'IA non produce nuovi pregiudizi e stereotipi: piuttosto amplifica quanto già esistente nel contesto sociale (Barocas e Selbst, 2016), in quello scambio "onlife" i cui limiti non esistono più. A livello sistemico i pregiudizi sono integrati nei processi automatizzati che governano molte delle nostre esperienze onlife. Gli algoritmi che determinano quali notizie vediamo, quali annunci pubblicitari riceviamo o quali contenuti ci vengono raccomandati, sono tutti influenzati da bias, sia espliciti sia impliciti. Secondo Noble (2018), questi sistemi possono

produrre un nuovo tipo di sorveglianza (Zuboff, 2019) e rafforzare le disuguaglianze esistenti, come nel caso degli algoritmi usati nella ricerca di notizie e/o prodotti che favoriscono contenuti discriminatori o sessisti. Si pensi, ad esempio, a studi come quello di Buolamwini e Gebru (2018), i quali hanno dimostrato che i sistemi di riconoscimento facciale tendono a essere meno accurati nel riconoscere i volti delle persone di altre culture rispetto a quelli con caratteristiche occidentali. Tali pregiudizi, quindi, possono indurre a prendere decisioni discriminatorie (Friedman e Nissenbaum, 1996). D'altra parte, «il sodalizio tra strutture sociali e strutture cognitive, che rappresenta “una delle più solide garanzie di dominio sociale”, dà luogo a quel potere simbolico che Bourdieu definisce come il “potere di costituire il dato nell’enunciato, di agire sul mondo agendo sulla rappresentazione del mondo» (Bourdieu e Wacquant, 1992:123 citato in Paolucci, 2002). Se applichiamo, dunque, questo impianto ai pregiudizi nell'IA, comprendiamo che essi derivano principalmente dai dati su cui i modelli (sociali e culturali) vengono addestrati e dalle scelte progettuali fatte dagli sviluppatori. Poiché i dati riflettono le disuguaglianze e i pregiudizi esistenti nella società, gli algoritmi di IA possono replicare e persino amplificare tali pregiudizi. Di conseguenza, da un punto di vista sociologico, risulta cruciale riflettere sul ruolo dell'individuo nella società come co-costruttore di conoscenza *sui generis*, proprio perché la stessa IA è “addestrata” da persone in relazione, le quali, costruendo una propria identità sociale, si fanno anche portatrici di possibili mitigazioni rispetto a bias/prejudizi propri dell'IA. L'IA ha, tuttavia, applicazioni positive e molto promettenti che vanno ugualmente sottolineate, in un bilanciamento doveroso fra rischi e opportunità. Proprio nel campo della giustizia sociale, essa ha il potenziale per promuovere l'equità e l'inclusione: l'analisi dei big data potrebbe essere impiegata per identificare e affrontare le

disuguaglianze sistemiche attraverso l'analisi dei grandi *dataset* in modo da rivelare *pattern* di discriminazione nelle pratiche di assunzione, nei prestiti bancari o nei processi giudiziari, contribuendo a rendere più trasparenti questi sistemi e a favorire politiche più eque (Passi e Barocas, 2019). I sistemi algoritmici intelligenti stanno anche trasformando il settore dell'assistenza sociale, migliorando l'autonomia e qualità della vita di persone anziane o con disabilità (Ferioli, 2019; Becchimanzi, 2022) e il settore dell'istruzione, dove permettono di personalizzare l'apprendimento e migliorare l'accessibilità del sapere (Luckin *et al.*, 2016). In coerenza, l'IA potrebbe aiutare a rendere più "democratico" l'accesso all'istruzione in quei contesti dove esso è limitato (aree rurali, aree interne o regioni affette da conflitti). Superare, o quanto meno governare, allora i pregiudizi nell'IA richiede un approccio multidimensionale che coinvolga la progettazione etica, la diversificazione dei dati e delle informazioni nonché il monitoraggio continuo anche da parte delle agenzie educative che, in accordo con l'acquisizione di un pensiero critico forte, hanno il dovere anche di far esercitare la partecipazione attiva degli educandi. Nell'"onlife" l'IA ha il potenziale per trasformare radicalmente la società, ma è fondamentale affrontare i pregiudizi intrinseci e valorizzarne le infinite opportunità. Comprendere le radici e le dinamiche *onlife* di questi pregiudizi risulta fondamentale per sviluppare strategie efficaci per mitigarli: tramite l'adozione di pratiche etiche, la promozione della *Digital Literacy* e la diversificazione nei processi di sviluppo tecnologico, invero, può essere possibile lavorare per una sintesi fra reale e virtuale più giusta, democratica ed inclusiva.

5. È possibile parlare di “inclusione sociale e digitale”?

L'IA, come brevemente descritto, rappresenta un potente strumento che, se adoperata in modo consapevole, critico e responsabile, può promuovere anche l'inclusione sociale. Quest'ultima è l'insieme delle politiche, pratiche e dei processi attraverso i quali individui e gruppi, spesso marginalizzati o esclusi, vengono “inclusi” pienamente nella società, acquisendo diritti, opportunità e risorse necessarie per partecipare attivamente alla vita economica, politica e culturale. Si tratta di una condizione multifattoriale strettamente connessa con il concetto di cittadinanza (Sen, 1999) che si riferisce, propriamente, all'esercizio dei diritti e delle responsabilità che legano l'individuo alla comunità. Non si tratta solo di integrazione economica ma anche e soprattutto del riconoscimento e del rispetto delle differenze culturali e identitarie (Silver, 1994; Bauman, 2000; Giddens, 2009). Applicando tale costrutto alla dimensione “onlife”, allora, l'IA potrebbe rappresentare quella possibilità di affrontare (almeno in parte) il problema delle disuguaglianze e delle barriere sistemiche che escludono gruppi emarginati dalla piena partecipazione alla società. L'inclusione digitale è, cioè, un aspetto essenziale dell'inclusione sociale nell'era tecnologica. Vale la pena ritornare, infatti, al momento pandemico più acuto che, nel nostro Paese, ha comportato molte sfide e criticità. Si pensi all'accesso alla rete che, in molti contesti geografici – soprattutto delle aree interne (Helsper, 2012) – non sempre è stato garantito unitamente alla difficoltà di poter usufruire della didattica a distanza per quei/quelle ragazzi/e stranieri e residenti in nuclei familiari numerosi o in condizioni abitative di sovraffollamento. Le famiglie – soprattutto quelle con minori a carico e gli stessi minori – con bassi livelli di istruzione, reddito, o che appartengono a gruppi socialmente esposti all'emarginazione, sono spesso svantaggiate

non solo in termini di accesso, ma anche nella capacità di utilizzare efficacemente le tecnologie digitali. Il *digital divide*, nelle forme sopra descritte di “esclusione sociale e digitale”, non si limita, quindi, alla semplice mancanza di accesso a internet o dispositivi tecnologici, ma comprende una serie di dimensioni più complesse e fra loro integrate, come le competenze digitali, la qualità dell’accesso, e l’uso significativo delle tecnologie (van Dijk, 2020). Un rischio tangibile che dovrebbe considerare il tema dell’inclusione sociale nella sua estensione digitale vista non solo come una questione tecnica di forniture e di accesso a internet, ivi compreso il sostegno economico per le dotazioni informatiche, ma come vero e proprio processo di *empowerment* sociale che riguarda anche l’alfabetizzazione digitale, la promozione delle competenze digitali e l’accesso a contenuti rilevanti e significativi. Come sottolineato da Selwyn (2004), l’inclusione (qui sociale e digitale) deve essere intesa come una condizione necessaria per la partecipazione sociale e civica in una società sempre più tecnologica, nell’esercizio della cittadinanza *onlife*. Ciò implica che le politiche pubbliche devono mirare non solo a ridurre il divario digitale, ma anche a promuovere l’uso critico e consapevole delle tecnologie informatiche. Il rispetto deve essere, quindi, per i diritti umani e democratici insieme alla possibilità di incrementare la *human capacity* in termini di promozione della partecipazione attiva di tutti i cittadini alla vita sociale, politica ed economica, riconoscendo e valorizzando la diversità culturale, linguistica e di genere. Riecheggia, in tal senso, l’approccio delle *capabilities* di Sen (1999) poiché si propone di valutare l’inclusione sociale e digitale non solo riguardo l’accesso e la disponibilità di risorse materiali, ma anche rispetto alle capacità effettive delle persone di agire e partecipare alla società tecnologica *tout court*. Il digitale può essere visto come un potenziale facilitatore di capacità. Per tutti questi motivi, di certo non esaustivi della questione, sarebbe utile

lavorare su più fronti che consentano democraticamente di potersi emancipare con e attraverso l'IA: le politiche pubbliche, i decisori politici, i corpi intermedi, le agenzie educative e di socializzazione e tutti noi dovremmo impegnarci seriamente, considerando le complesse sfide che attendono l'umano, per poter creare un mondo *onlife* che sia più accessibile ed equo.

Conclusione

La *Digital Literacy*, alla luce di quanto brevemente esposto, è una competenza essenziale per navigare il complesso ecosistema tecnologico odierno, dove l'IA gioca un ruolo sempre più centrale. La diffusione, l'impiego e il miglioramento delle tecnologie su base IA, infatti, impongono nuove sfide e criticità ma anche notevoli opportunità per coloro che sono in grado di comprenderle e utilizzarle con consapevolezza. Ecco perché, per affrontare al meglio questi cambiamenti, è cruciale promuovere un'alfabetizzazione digitale che vada oltre le competenze tecniche di base, orientandosi verso una comprensione critica delle tecnologie e delle dinamiche di potere che le sottendono. Infatti, quando le tecnologie, IA compresa, vengono utilizzate da persone competenti, esse sono in grado di migliorare effettivamente la qualità di vita, promuovendo l'inclusione attiva e responsabile (D'Ambrosio e Boriati, 2023:26) degli attori sociali. Una possibile risposta istituzionale, che aumenti le competenze digitali di ciascuno, potrebbe essere quella di inserire nei curricula scolastici e accademici percorsi di tipo socio-pedagogico trasversali che lavorino sulla consapevolezza dei discenti nell'utilizzo delle tecnologie digitali *tout court*, IA compresa. Tali strategie si pongono

in continuità con quanto già previsto a livello nazionale¹ e sovranazionale², le quali offrono orientamenti non solo di tipo tecnico ma anche etico nell'utilizzo dell'IA. Soltanto attraverso utenti-cittadini consapevoli e ben informati, dunque, è possibile sperare di costruire un futuro digitale inclusivo, etico, giusto e sicuro, in cui l'IA non sia soltanto uno strumento di efficienza, ma anche un motore di progresso etico e sociale che contribuisca al perseguimento del bene(-essere) comune. L'IA, concludendo, rappresenta un'opportunità straordinaria per potenziare le capacità umane, ma senza un'adeguata *Digital Literacy* – che sia in grado di promuovere l'utilizzo della tecnologia in modo competente, critico e riflessivo, permettendo altresì di partecipare attivamente al mondo sociale, sempre più digitalizzato e connesso –, essa rischia di ampliare le disuguaglianze sociali e di erodere i confini della libertà di pensiero e di parola. Può mettere a repentaglio, in altri termini, la capacità di adempiere ai propri diritti di cittadinanza, tanto *online* quanto *off-line*.

Bibliografia

Arrow, K. (1973). The theory of discrimination. *Discrimination, Labor Markets*, 3(10), 3-33.

Baker, R. S. & Hawn, M. A. (2022). Algorithmic Bias in Education. *International Journal of Artificial Intelligence and Education*, 32, 1052-92.

¹ Si veda, tra gli altri, la strategia nazionale sull'utilizzo dell'IA. Cfr: <https://www.agid.gov.it/it/agenzia/stampa-e-comunicazione/notizie/2023/11/09/intelligenza-artificiale-presentato-il-comitato-coordinamento-laggiornamento> (ultimo accesso: 13/02/2025).

² Si veda, a tal proposito, il link <https://digital-strategy.ec.europa.eu/en/policies/digital-principles> (ultimo accesso: 13/02/2025).

Barocas, S. & Selbst, A. D. (2016). Big Data's Disparate Impact, *California Law Review*, 104(3), 671-732.

Bauman, Z. (2000). *Liquid Modernity*. Cambridge: Polity Press.

Becchimanzi, C. (2022). Sfide di progettazione per l'accettabilità nell'interazione uomo-robot: framework e strumenti per migliorare l'esperienza dell'utente e l'interazione con le tecnologie emergenti, *Rivista Italiana di Ergonomia*, 267-72.

Bechmann, A. & Bowker, G. C. (2019). Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media. *Big Data & Society*, 6(1), 1-11.

Beer, D. (2017). The social power of algorithms, *Information, Communication & Society*, 20(1), 1-13.

Berger, P. T. & Luckmann, T. (1966). *The Social Construction of Reality*. New York: Anchor Books.

Binns, R. (2018), Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR 81, 149-59.

Bloomfield, B. P. (1987). The culture of artificial intelligence. In B. P. Bloomfield (Eds.), *The question of artificial intelligence. Philosophical and sociological perspectives*. New York: Croom Helm. 59-105.

Bourdieu, P. (1975). The specificity of the scientific field and the social conditions of the progress of reason, *Social Science Information*, 14(6), 19-47.

Bourdieu, P. (1998). *Il dominio maschile*. Milano: Feltrinelli.

Bourdieu, P. & Wacquant, L. (1992). *Réponses. Pour une anthropologie reflexive*. Paris: Seuil.

Broussard M. (2018). *Artificial unintelligence: How computers misunderstand the world*. Cambridge, MA: The MIT Press.

Buolamwini, J. & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 77-91.

Castells, M. (1996). *The Rise of the Network Society*. Hoboken: Wiley-Blackwell.

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M. & Floridi, L. (2018). Artificial Intelligence and the “Good Society”: The US, EU, and UK Approach. *Science and Engineering Ethics*, 24(2), 505-28.

Centorrino, M. & Romeo, A. (2023), *Media digitali. La costruzione delle relazioni sociali*. Milano: FrancoAngeli.

Commission of the European Communities. (2003). *Joint Report on Social Inclusion summarizing the results of the examination of the National Action Plans for Social Inclusion (2003-2005)*. Brussels. COM 773.

Commissione Europea (2018). *Digital Competence Framework for Citizens*. Brussels.

Commissione Europea (2020). *On artificial intelligence - a European approach to excellence and trust*. White Paper. Brussels.

Cruz, T. M. (2020). Perils of data-driven equity: safety-net care and big data's elusive grasp on health inequality. *Big Data & Society*, 7(1). <https://doi.org/10.1177/2053951720928097>.

Cuomo, S., Biagini, G. & Ranieri, M. (2022). Artificial Intelligence Literacy, che cos'è e come promuoverla. Dall'analisi della letteratura ad una proposta di Framework, *Media Education*, 13(2), 161-72. doi: 10.36253/me-13374.

D'Ambrosio M. & Boriati D. (2023). Digital Literacy, Technology Education and Lifelong Learning for Elderly: Towards Policies for a Digital Social Innovation Welfare. *Italian Journal of Sociology of Education*, 15(2), 21-36.

De Luca, C., Domenici, G. & Spadafora, G. (2023). *Per una inclusione sostenibile. La prospettiva di un nuovo paradigma educativo*. Roma; Editoriale Anicia.

Edelmann, A., Wolff, T., Montagne, D. & Bail, C. A. (2020). Computational social science and sociology. *Annual Review of Sociology*, 46, 61-81. <https://doi.org/10.1146/annurev-soc-121919-054621>.

Eshet-Alkalai, Y. (2004). Digital Literacy: A Conceptual Framework for Survival Skills in the Digital Era. *Journal of Educational Multimedia and Hypermedia*, 13(1), 93-106.

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

Facione, P. A. (2011). Critical Thinking: What It Is and Why It Counts, *Insight Assessment*, 2013, 1-28. <https://www.law.uh.edu/blkely/advocacy-survey/Critical%20Thinking%20Skills.pdf>.

Feroli, E. (2019). L'intelligenza artificiale nei servizi sociali e sanitari: una nuova sfida al ruolo delle istituzioni pubbliche nel welfare italiano? *BioLaw Journal - Rivista di BioDiritto*, 1, 163-75.

Fleck, J. (1987). Development and establishment of artificial intelligence, In B.P. Bloomfield (Eds.), *The question of artificial intelligence: Philosophical and sociological perspectives*. New York: Croom Helm, 106-64.

Floridi, L. (2015). *The Onlife Manifesto: Being Human in a Hyperconnected Era*. Cham; Heidelberg; New York; Dordrecht; London: Springer Cham. <https://link.springer.com/content/pdf/10.1007/978-3-319-04093-6.pdf>.

Floridi, L. (2018). *The Ethics of Artificial Intelligence*. Berlin: Springer International Publishing.

Floridi, L. (2019). *The Logic of Information: A Theory of Philosophy as Conceptual Design*. Oxford: Oxford University Press.

Forsythe, D. E. (1993a). The construction of work in artificial intelligence. *Science, Technology, & Human Values*, 18(4), 460-79.

Forsythe, D. E. (1993b). Engineering knowledge: The construction of knowledge in artificial intelligence. *Social Studies of Science*, 23(3), 445-77.

Franzoni, V. (2023). Gender Differences and Bias in Artificial Intelligence. In Vallverdú, J. (2023). *Gender in AI and Robotics. The Gender Challenges from an Interdisciplinary Perspective*, Intelligent Systems Reference Library Volume 235. Springer Cham. 27-44.

Friedman, B. & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems*, 14(3), 330-47.

Giddens, A. (2009). *Sociology*. Cambridge: Polity Press.

Gilster, P. (1997). *Digital Literacy*. Wiley Computer Publishing.

Golder, S. A. & Macy, M. W. (2014). Digital footprints: Opportunities and challenges for online social research. *Annual Review of Sociology*, 40, 129-52.

Hargittai, E. (2002). Second-Level Digital Divide: Differences in People's Online Skills. *First Monday*, 7(4).

Helsper, E. J. (2012). A Corresponding Fields Model for the Links Between Social and Digital Exclusion. *Communication Theory*, 22(4), 403-26.

Hildebrandt, M. (2015), Dualism is Dead. Long Live Plurality (Instead of Duality). In Floridi, L. (Eds.). *The Onlife Manifesto*, Berlin: Springer, Cham.

Hofeditz, L., Mirbabaie, M., Luther, A., Mauth, R. & Rentemeister, I. (2022). Ethics guidelines for using AI-based algorithms in recruiting: Learnings from a systematic literature review. *Proceedings of the 55th Hawaii International Conference on System Sciences*, 2022, 145-54.

Hughes, T. P. (1987), The Evolution of Large Technological Systems. In Bijker W., Hughes T. P., Pinch T., *The Social Construction of Technological Systems*. Cambridge, MA: MIT Press. 45-76.

Istituto Nazionale di Statistica (2023). *Cittadini e Competenze Digitali*. Roma.

Kahneman, D., Slovic, S. P., Slovic, P. & Tversky, A. (1982). *Judgment under uncertainty: heuristics and biases*. Cambridge: Cambridge University Press.

Köchling, A. & Wehner, M. C. (2020). Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, 13(3), 795-848.

Leavy, S. (2018). Gender Bias in Artificial Intelligence: The Need for Diversity and Gender Theory in Machine Learning. 2018 ACM/IEEE 1st International Workshop on Gender Equality in Software Engineering, 14-6.

Lee, N. T. (2018). Detecting racial bias in algorithms and machine learning. *Journal of Information, Communication and Ethics in Society*, 16(3), 252-60.

Leitgöb, H., Prandner, D. & Wolbring, T. (2023). Editorial: Big data and machine learning in sociology. *Frontiers of Sociology*, 8.

Liu, Z. (2021). Sociological perspectives on artificial intelligence: A typological reading. *Sociology Compass*, 15, 1-13.

Lloyd, K. (2018). Bias Amplification in Artificial Intelligence Systems, arXiv preprint arXiv:1809.07842. <https://doi.org/10.48550/arXiv.1809.07842>.

Long, D. & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. Proceedings of the 2020 CHI Conference on Human Factors and Computing Systems, 1-16.

Luckin, R., Holmes, W., Griffiths, M. & Forcier, L. B. (2016). *Intelligence Unleashed: An Argument for AI in Education*, Pearson.

McLuhan, M. (1967). *Gli strumenti del comunicare*. Milano: Il Saggiatore.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. & Floridi, L. (2016). The ethics of algorithms: mapping the debate. *Big Data & Society*, 3(2), 1-21.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York: University Press.

Norman, D. A. & Draper, S. W. (1986). *User-Centered System Design: New Perspectives on Human-Computer Interaction*. Mahwah: Lawrence Erlbaum Associates.

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), 447-53.

Paolucci, G. (2002). L'impegno della ragione sociologica contro le maschere del dominio. *Quaderni di Sociologia*, 29, 151-61.

Parikh R. B., Teeple S. & Navathe A. S. (2019). Addressing Bias in Artificial Intelligence in Health Care, *JAMA*, 322(24), 2377-78.

Passi, S. & Barocas, S. (2019). Problem formulation and fairness, Proceedings of the ACM conference on fairness, accountability, and transparency, New York. 39-48.

Prensky, M. (2001). Digital Natives Digital Immigrants, On the Horizon. *MCB University Press*, 9(5).

Prensky, M. (2009). H. Sapiens Digital: From Digital Immigrants and Digital Natives to Digital Wisdom. *Innovate: Journal of Online Education*, 5(3). <https://www.learntechlib.org/p/104264/>.

Prensky, M. (2010). H. Sapiens Digitale: dagli Immigrati digitali e nativi digitali alla saggezza digitale. *TD-Tecnologie Didattiche*, 50, 17-24.

Rahwan, I. (2018). Society-in-the-loop programming the algorithmic social contract. *Ethics Inf Technol*, 20(1), 5-14.

Rogers, E. (2003). *Diffusion of Innovation*. New York: Free Press.

Schwartz, R. D. (1989). Artificial intelligence as a sociological phenomenon. *Canadian Journal of Sociology*, 14(2), 179-220.

Selwyn, N. (2004). Reconsidering Political and Popular Understandings of the Digital Divide. *New Media & Society*, 6(3), 341-62.

Sen, A. (1999). *Development as Freedom*. Oxford: Oxford University Press.

Silver, H. (1994). Social Exclusion and Social Solidarity: Three Paradigms. *International Labour Review*, 133(5-6), 531-78.

Silverstone, R. & Hirsch, E. (1992). *Consuming Technologies. Media and Information in Domestic Spaces*. London: Routledge.

Tufekci, Z. (2014). Big questions for social media big data: representativeness, validity and other methodological pitfalls, Proceedings of the Eight International AAAI Conference on Weblogs and Social Media, 505-14.

van Dijk, J. A. (2020). *The Digital Divide*. Cambridge: Polity Press.

van Dijck, J. A., Poell, T. & de Waal, M. (2019). *Platform society. Valori pubblici e società connessa*. Milano: Guerini Scientifica.

Varsha, P. S. (2023), How can we manage biases in artificial intelligence systems – A systematic literature review, *International Journal of Information Management Data Insights*, 3(1).

Walsh, T., Bell, G., Brundage, M., Griffin, P. & Yampolskiy, R. V. (2020). Artificial Intelligence: Opportunities and Risks. *AI & Society*, 35(4), 625-40.

Winner, L. (2020). *The Whale and the Reactor: A Search for Limits in an Age of High Technology*. Chicago: The University of Chicago Press.

Yeh, S. C., Wu, A. W., Yu, H. C., Wu, H. C., Kuo, Y. P. & Chen, P. X. (2021). Public perception of artificial intelligence and its connections to the sustainable development goals. *Sustainability*, 13(16), 9165.

Zheng, L. (2021). Sociological perspectives on artificial intelligence: A typological reading. *Sociology Compass*, 15(3), 1-13.

Zuboff, S. (2019). *The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power*. New York: Public Affairs.

3. RIATTIVARE LA RIFLESSIVITÀ: PER UN MODELLO ETICO-CRITICO DI EDUCAZIONE DIGITALE

di Giuseppe De Ruvo*

Abstract: L'articolo propone un modello di educazione digitale critica (CDL) per affrontare le dinamiche di potere del capitalismo digitale, evidenziando i limiti dell'attuale digital literacy (DL). La DL, tradizionalmente orientata all'acquisizione di competenze tecniche per il lavoro, è criticata per il suo approccio astratto e ideologico, che ignora il contesto socioeconomico. La CDL, invece, integra l'analisi critica delle relazioni tra tecnologie e potere, superando la logica adattiva della DL. Questo approccio trova fondamento nella pedagogia critica di Paulo Freire, che distingue tra educazione "depositaria" e "problematizzante". Mentre la prima trasmette nozioni prive di contesto, la seconda stimola una comprensione critica, contestualizzando i fenomeni e smascherandone le basi ideologiche. Applicata al digitale, la CDL utilizza questa prospettiva per decostruire narrazioni ideologiche, come il mito dell'oggettività algoritmica, e formare cittadini consapevoli e attivi. L'articolo argomenta che la CDL può fungere da catalizzatore per una teoria critica del capitalismo digitale, riattivando la riflessività dei soggetti e favorendo la costruzione di utopie emancipative. In un'epoca dominata dal mito dell'oggettività algoritmica e dalla compressione dello spazio riflessivo, l'educazione critica diventa essenziale per ripensare le relazioni tra tecnologia, società e potere.

Parole chiave: AI, Critical digital literacy, Riflessività, Teoria Critica, Utopia.

* Università Vita-Salute San Raffaele, giuseppederuvo1@gmail.com.

Accettato Dicembre 2024 - Pubblicato Aprile 2025.

Abstract: The article proposes a model of critical digital education (CDL) to address the power dynamics of digital capitalism, highlighting the limitations of current digital literacy (DL). DL, traditionally geared towards the acquisition of technical skills for work, is criticised for its abstract and ideological approach, which ignores the socio-economic context. CDL, on the other hand, integrates critical analysis of the relationship between technologies and power, overcoming the adaptive logic of DL. This approach is grounded in Paulo Freire's critical pedagogy, which distinguishes between 'depository' and 'problematizing' education. While the former conveys notions without context, the latter stimulates critical understanding, contextualising phenomena and unmasking their ideological foundations. Applied to the digital, CDL uses this perspective to deconstruct ideological narratives, such as the myth of algorithmic objectivity, and form conscious and active citizens. The article argues that CDL can act as a catalyst for a critical theory of digital capitalism, reactivating the reflexivity of subjects and fostering the construction of emancipatory utopias. In an era dominated by the myth of algorithmic objectivity and the compression of reflexive space, critical education becomes essential for rethinking the relationships between technology, society, and power.

Keywords: AI, Critical digital literacy, Reflexivity, Critical theory, Utopia.

1. Automatizzare le esistenze. Il potere del capitalismo digitale

Negli ultimi anni, la letteratura ha iniziato a occuparsi in maniera sistematica delle pratiche di potere connesse alle nuove tecnologie digitali. In una prima fase, infatti, queste ricerche si erano concentrate soprattutto sull'analisi delle singole forme di potere o discriminazione scaturite dall'adozione intensiva delle nuove tecnologie, mostrando – ad esempio – in che modo avvenisse l'estrazione e la manipolazione (non consensuale) dei dati degli utenti (Van Dijck, 2014), il modo in cui gli algoritmi riproducono

stereotipi presenti nella società (dunque rafforzandoli) (Noble, 2018) e il rischio corso dalle istituzioni – governative e non – nell'affidarsi a pratiche algoritmiche *black-boxed* su cui non hanno un controllo né tecnico né epistemico (Pasquale, 2015). E tuttavia, a partire dagli studi di autori come Srnicek (2016) e Zuboff (2019), siffatte pratiche di potere hanno iniziato a essere analizzate in un contesto più ampio, ovvero quello caratterizzato dall'imporsi, a livello economico, del capitalismo digitale. Insomma, per comprendere la portata del potere che transita nei *media digitali*, è diventato necessario riferirsi alle pratiche di ristrutturazione capitalistiche e alle loro conseguenze in termini di produzione di soggettività, attraverso un impegno critico "inclusivo" che integra approcci tipici del marxismo (Fuchs, 2021), della Scuola di Francoforte (Delanty & Harris, 2021), degli studi sulla governamentalità di stampo foucaultiano (Berns & Rouvroy, 2013) e della teoria dei sistemi socio-tecnici (STS). I risultati a cui questi studi sono giunti, pur nelle loro differenze, sono abbastanza coerenti tra di loro (almeno della *pars destruens*) e hanno condotto a due risultati principali:

- a) Il capitalismo digitale, lungi dal costituire una rottura rivoluzionaria nella storia del capitalismo, si presenta in realtà come una sua razionalizzazione in molteplici ambiti, che vanno dall'organizzazione aziendale alla distribuzione delle merci, passando per la realizzazione del valore di scambio e lo sfruttamento del lavoro (Pfeiffer, 2022; Rouvroy, 2016; Langley & Leyshon, 2016);
- b) Il capitalismo digitale ha una fortissima dimensione cognitiva e ideologica (Fumagalli, 2007), nella misura in cui esso necessita di "soggetti collaborativi" (Pierazzuoli, 2022) che seguano le prescrizioni e le raccomandazioni algoritmiche, *conditio sine qua non* per fluidificare il sistema e raggiungere la razionalizzazione sopra descritta.

Ovviamente, queste due dimensioni non possono essere tenute astrattamente separate: la dimensione “cognitivo-ideologica” e quella “economico-industriale” si sostengono e integrano vicendevolmente, dunque non è possibile pensare questa distinzione – puramente analitica e operativa – in termini essenzialistici, di fatto riproducendo un dualismo tra struttura economica e sovrastruttura ideologica. Tuttavia, dal momento che l’obiettivo del nostro articolo è quello di tematizzare un’educazione digitale che possa accompagnare e favorire lo sviluppo di un pensiero e di una prassi critica nei confronti del capitalismo digitale, è utile – seppur brevemente – concentrarci sulla sua dimensione ideologico-cognitiva. Per comprendere l’agire ideologico del capitalismo digitale, è fondamentale tenere a mente come, allo stato dell’arte, gli algoritmi che governano la nostra esistenza online scontino ancora profonde limitazioni. Queste non dipendono da questioni contingenti, quanto dal fatto che gli attuali algoritmi, quando si trovano ad affrontare scenari complessi come quelli che riguardano le situazioni umane, tendono a sviluppare una “irriducibile tendenza all’errore” (Possati, 2021: 80) che dipende principalmente da due fattori, i quali costituiscono il cosiddetto “problema dell’implementazione” (Finn, 2017). In primo luogo, gli algoritmi – per definizione – possono avere a che fare solo con un numero finito di dati – o comunque con un numero di dati che gli permetta di svolgere le loro operazioni in un tempo ragionevole senza saturare la memoria del calcolatore (Angius, 2023). Dunque, non è assolutamente certo che essi dispongano di tutti i dati necessari per affrontare una particolare situazione, e non è neppure detto che gli sviluppatori siano in grado di rimediare, dato che decidere quali dati debbano entrare a far parte di un *data-set* di partenza o d’allenamento è, allo stato dell’arte, un problema indecidibile (Ben David *et al.*, 2019). In secondo luogo, le predizioni algoritmiche – già di per sé basate su modelli probabilistici –

scontano anche inevitabili approssimazioni, che dipendono dal processo tecnico-ingegneristico di indicizzazione delle variabili – necessario per evitare l’esplosione combinatoria e garantire l’efficienza computazionale dell’algoritmo. Infatti, la traduzione di uno stato di cose in un sistema binario di 0 e 1 implica “l’approssimazione nel discreto del modello dato sul continuo” (Zellini, 2010: 334), con conseguente approssimazione anche nel risultato. Ovviamente, ciò non significa che siffatte predizioni non siano *utili*. Al contrario, come è stato notato da più autori, esse agiscono come degli efficaci meccanismi di riduzione della complessità che sono in grado di sgravare i soggetti che navigano online dal “fardello della scelta” (Cohn, 2019), organizzando la loro esperienza *online* e rendendola più fluida (Csikszentmihalyi, 2008) e personalizzata. Questa riduzione della complessità portata dagli algoritmi è affiancata, da parte delle grandi aziende digitali, da una retorica ideologica circa l’onnipotenza dei sistemi algoritmici. Come scrive David Beer (2019: 25), in uno studio in cui analizza le retoriche di 34 aziende digitali, le piattaforme tendono a descrivere i loro dispositivi come “onnipotenti”. Inconoscibili e totalizzanti nel loro sguardo. Nulla può sfuggire questa vista profetica e rafforzata tecnologicamente”. Gli algoritmi delle Big Tech vengono insomma presentati come dotati di una “naturale tendenza all’oggettività” (Illis, Petit, Jarret, 2013, 37) in virtù della quale essi non sarebbero solo in grado di ridurre la complessità della vita degli utenti, ma addirittura di realizzare il “sogno della serendipità sistemica” (Berns & Rouvroy, 2013: 19) e di “offrire una conoscenza del mondo e di sé più vera, più ricca e precisa” (Fisher, 2022: 11). Lo sviluppo di questa narrazione ideologica è cruciale nell’economia del capitalismo digitale, giacché grazie a essa – e in particolare grazie al fatto che essa “media non solo la relazione degli utenti con il loro ambiente socioculturale ma anche con loro stessi” (Karakayali *et al.*, 2018: 6) – i soggetti si identificano in maniera

crescente con il loro doppio digitale e con le protensioni generate algebricamente (Stiegler, 2017), così comportandosi *come se* le previsioni degli algoritmi fossero prive dei limiti che abbiamo descritto in precedenza. In sintesi: dinnanzi ai limiti tecnici che affliggono i loro algoritmi, le aziende digitali hanno attuato una svolta performativa, in virtù della quale non sono più (solo) gli algoritmi ad adattarsi ai soggetti, giacché – attraverso narrazioni ideologiche come quella sopradescritta – si tenta di far sì che siano questi ultimi ad adattarsi agli algoritmi, ovvero a soggettivizzarsi sulla base delle normatività da essi generate (Lupton, 2016; Cheney-Lippold, 2017; Scaltsas, 2017). Come rilevato da molti autori, queste pratiche mirano ad aumentare il tasso di successo delle previsioni algebriche, facendo sì che i soggetti, identificandosi con esse, avverino in misura maggiore le prescrizioni degli algoritmi. L'obiettivo è quello di aumentare i ricavi da parte degli inserzionisti, le cui spese in *advertising* sono direttamente proporzionali al tasso di conversione e realizzazione del valore della merce reso possibile dalle piattaforme (Srnicsek, 2016; Langley & Leyshon, 2016; Barry, 2019; Zuboff, 2019). Ma cosa avviene agli utenti quando entrano in queste dinamiche di potere, attraverso le quali essi vengono soggettivizzati sulla base delle normatività algebriche e resi soggetti "accomodanti"?

Per usare l'espressione di Pierre Bourdieu (1990), recuperata da diversi studiosi di *media studies* (Romele, 2023; Hristova, 2023), l'azione combinata di ideologia dell'oggettività algebrica, accelerazione della vita digitale (Wajcman, 2014) e *digital nudging* (Yeung, 2017) genera un "*digital habitus*", ovvero una postura in virtù della quale i soggetti vivono la loro esistenza digitale in maniera "irreflessa". Fondamentalmente, per il capitalismo digitale è cruciale "evitare" che i soggetti mettano in campo pratiche riflessive, perché – da un lato – queste potrebbero portarli a criticare le sue pratiche sulla base di standard etico-

morali e – dall’altro – perché esse potrebbero far sì che i soggetti, riflettendo su loro stessi, rompano la catena causale algoritmica, riducendo il tasso di successo delle predizioni degli algoritmi (Fisher, 2022). In questo articolo, cercheremo dunque di mostrare come un’educazione digitale critica possa impattare positivamente su questa congiuntura, proprio nella misura in cui essa, stimolando le capacità riflessive dei soggetti, renderebbe più complesso per le piattaforme instillare in essi dei “deficit strutturali di riflessività” (Celikates, 2018: 175). Per farlo, nel prossimo paragrafo (§2) analizzeremo criticamente l’attuale letteratura sulla *digital literacy*, per poi elaborare un modello di educazione digitale critico e orientato all’emancipazione (§3). In conclusione (§4), si tematizzerà il ruolo e l’importanza di siffatto modello all’interno di una teoria critica del capitalismo digitale.

2. Una prospettiva critica sulla digital literacy

Dato il rapido sviluppo dell’AI e delle tecnologie a essa legate, diversi studiosi hanno tematizzato e messo a punto dei percorsi formativi che mirano a inserire l’alfabetizzazione digitale nei *curricula* scolastici, in modo da formare cittadini pronti alle sfide di domani. E tuttavia, come si vedrà nel corso di questo paragrafo, queste proposte scontano dei limiti importanti, che impediscono ai discenti di appropriarsi riflessivamente delle nuove tecnologie e, dunque, di trasformarsi in cittadini effettivamente attivi e consapevoli del mondo digitale. Come si vedrà, la *digital literacy* (DL) sconta tre limiti principali, messi in luce dalla *critical digital literacy* (CDL), ovvero l’astrattezza, l’eccessiva concentrazione sull’acquisizione di abilità e la sua natura ideologica. Per comprendere questi problemi, è utile iniziare ponendo una domanda. Ovvero: perché la DL è considerata centrale? La

letteratura risponde in maniera sostanzialmente unanime: la *digital literacy* è necessaria affinché i discenti “ottengano un vantaggio competitivo nel lavoro [...]”. Nell’attuale epoca digitale, l’emergere della società della conoscenza implica che ogni cittadino debba essere alfabetizzato digitalmente per essere in grado di accedere alle opportunità offerte dal mondo del lavoro” (Ng *et al.*, 2021: 1). La DL serve innanzitutto per “garantire impiegabilità e creare opportunità di carriera” (Su & Zhong, 2022: 2). Insomma, dato che “nel prossimo futuro i lavori saranno largamente legati all’AI” (Kandlhofer *et al.*, 2016: np), è cruciale che i giovani “inizino a imparare queste abilità di base sin dall’inizio” (Ibidem). Effettivamente, il legame tra alfabetizzazione digitale e mondo del lavoro ha una sua profondità storica. Come notano Leahy e Dolan (2010), l’espressione stessa “*digital literacy*” è nata negli anni Settanta in ambito aziendale, quando “si rese necessario per gli utenti diventare competenti nell’uso dei computer per quanto riguardava i compiti che dovevano svolgere nel loro lavoro” (Ng *et al.*, 2021: 3). Di conseguenza, data la sua centralità per il mondo del lavoro, la DL viene intesa come una forma di conoscenza applicativa, in tutto e per tutto paragonabile alla trasmissione delle tecniche di scrittura, lettura e calcolo (Ivi: 1). L’obiettivo è quello di rendere i discenti in grado di pensare “computazionalmente”, ovvero in grado di “risolvere problemi, costruire sistemi e comprendere il comportamento umano sulla base della *computer science*” (Su & Zhong, 2022: 3). Anche la conoscenza di base offerta agli studenti è orientata verso questo obiettivo. Secondo Kandlhofer e colleghi (2016), infatti, la DL riguarda 1) la capacità di descrivere sistemi automatici (*automata*), 2) lo studio degli agenti intelligenti e dei loro processi decisionali (*intelligent agents*), 3) la familiarità con grafici e strutture di dati (*graphs and data structures*), 4) la capacità di risolvere problemi in maniera soddisfacente attraverso algoritmi (*problem solving by search*), 5) la

capacità di scrivere piani d'azione per i programmi attraverso lo studio della logica (sia proposizionale sia predicativa) (*classic planning*) e, 6), la conoscenza dei principali modelli di apprendimento automatico (*machine learning*). Similmente, le “five big ideas” sottese alla *digital literacy* – tematizzate da Touretzky e colleghi (2019) – sono parte integrante dei diversi *curricula*, e riguardano 1) la consapevolezza che i computer percepiscono il mondo, 2) lo studio delle rappresentazioni del mondo degli agenti artificiali, 3) l'essere a conoscenza della capacità dei computer di imparare dai dati, 4) il fatto che gli agenti artificiali possono interagire con gli umani e 5) la questione dell'impatto sociale dell'AI. Il framework proposto da Kim e colleghi (2021) si concentra invece su cinque “cluster” di alfabetizzazione, che riguardano 1) definizione e tipi di AI, 2) risoluzione di problemi e ricerca, 3) ragionamento, 4) dati e *machine learning*, 5) applicazioni.

Infine, il *curriculum* messo a punto da Su e Zhong (2022) vede la successione di quattro moduli. Il primo è dedicato alla storia dell'AI (a cui vengono dedicati 30 minuti) e alla comprensione delle cinque grandi idee a essa sottese (Touretzky *et al.*, 2019); il secondo è dedicato al *machine learning*, il terzo alla *speech recognition*, mentre il quarto ha al centro lo studio dei *bias* e la loro mitigazione. Come ben riassunto da Ng e colleghi (2021: 10), il razionale sotteso a questi framework deriva dal fatto che “le persone devono essere equipaggiate cognitivamente per le prossime sfide tecnologiche nei loro posti di lavoro”. La DL deve dunque combinare “le idee della *data science*, del pensiero computazionale e della conoscenza multidisciplinare” con l'obiettivo, ben espresso da Kandlhofer e colleghi (2016), di rendere i giovani, intesi come lavoratori di domani, “fluenti nell'uso dell'AI”, ovvero in grado di “applicare metodi di risoluzione dei problemi a un alto livello d'astrazione” (Ibidem). Ovviamente, non è possibile negare l'importanza che le tecnologie digitali hanno (e avranno) nel mondo

del lavoro, dunque formare cittadini in grado di sfruttarle rimane una questione centrale. E tuttavia, la DL può essere criticata (e integrata) sotto un duplice punto di vista: il primo è legato allo statuto ontologico delle tecnologie digitali, il secondo – e conseguente – riguarda l’ambito oggettuale della *digital literacy*. Per quanto riguarda la prima questione, l’analisi dei diversi framework svolta in precedenza manifesta come questi programmi approccino l’AI e le tecnologie digitali da un punto di vista prevalentemente tecnico-ingegneristico, sostanzialmente *astruendo* dalle concrete relazioni che le tecnologie digitali intessono con il tessuto produttivo e socioeconomico dal quale sgorgano e che, a loro volta, influenzano. Come buona parte della letteratura sociologica e filosofica ha oramai sottolineato, infatti, l’AI e le tecnologie digitali non possono essere comprese se analizzate come puri fatti tecnici, giacché esse sono in realtà prodotte e composte da un sistema multistrato d’interazione, in cui fattori tecnici, economici, culturali e filosofici – spesso in contraddizione tra loro – s’intersecano, dando vita a quegli “assemblaggi sociotecnici” che siamo soliti chiamare “intelligenza artificiale” (Lindgren, 2023) o “algoritmi” (Gillespie, 2016). Queste considerazioni ci portano alla seconda questione, che riguarda l’ambito oggettuale della *digital literacy*. Se, infatti, le tecnologie digitali non possono essere considerate come dei meri oggetti tecnici, ma come degli assemblaggi sociotecnici che emergono e agiscono in un contesto più ampio, allora la DL non può limitarsi ai framework descritti in precedenza, nella misura in cui – per avere una chiara consapevolezza dell’impatto sociale delle nuove tecnologie, elemento sottolineato da tutti i *curricula* presi in esame – è fondamentale tematizzare il contesto socioeconomico in cui queste si sviluppano e che contribuiscono a disegnare. La DL dovrebbe dunque trascendere la mera dimensione applicativa e contenutistica orientata al mondo del lavoro, per trasformarsi a tutti gli effetti in una educazione

digitale critica, in grado di tematizzare le tecnologie digitali nel loro rapporto con la società. È prendendo le mosse da considerazioni di questo genere che, al fianco della *digital literacy*, si è sviluppata la *critical digital literacy* (CDL), la quale contesta i framework offerti dalla DL proprio in virtù della loro postura centrata sull'applicazione a fini lavorativi, della natura astratta di queste pratiche e, in fin dei conti, del loro essere ideologiche. Iniziamo dalla dimensione applicativa della DL. Come si è visto, essa deriva – anche storicamente – dal suo stretto rapporto col mondo del lavoro. Ora, come notano Špiranec e colleghi (2019), siffatto approccio pare rivolgersi solo a potenziali professionisti, che si troveranno ad avere a che fare attivamente con l'AI. E tuttavia, esistono soggetti che, nel corso della loro vita, avranno un rapporto soprattutto “passivo” con le nuove tecnologie digitali. In una parola: non per tutti l'AI e gli algoritmi saranno strumenti di lavoro. Anzi, molti cittadini avranno a che fare con queste tecnologie esclusivamente nella misura in cui esse, incarnate in algoritmi di raccomandazione o di altro genere, prenderanno decisioni *per loro o al loro posto* (Pasquale, 2015). Ed è proprio per tutelare queste fasce della popolazione che, secondo Špiranec e colleghi (2019: np), la CDL dovrebbe “non solo concentrarsi sui professionisti, ma anche [...] – e soprattutto – sui cittadini”. Da questo punto di vista, la CDL si impegna a tematizzare non solo la dimensione pratico-applicativa delle nuove tecnologie digitali, ma anche quella politico-epistemologica, con particolare attenzione alla scienza dei dati. La CDL, infatti, “vede i dati come un qualcosa di sempre problematico” (Ibidem) e, dunque, vuole “assicurarsi che gli studenti non imparino solo a fare analisi statistiche [...] ma che conoscano anche i presupposti filosofici del lavoro scientifico, l'organizzazione sociale delle scienze e l'influenza della società sulla produzione di conoscenza scientifica” (Ibidem). Combinare la conoscenza scientifica sottesa alle nuove tecnologie digitali con la

conoscenza del contesto socioeconomico in cui queste operano, dunque delle dinamiche ideologiche e di potere che le attraversano, è semplicemente essenziale per formare non solo professionisti, ma anche cittadini “proattivi”, ovvero in grado di pensare *criticamente* le nuove tecnologie. Infatti, come nota Victoria Carrington (2018: 72), l’assenza di CDL non avviene in un contesto scevro da dinamiche di potere.

L’assenza di impegno critico degli utenti, la differenza di potere tra utenti e aziende nell’uso dei dati, tra l’individuo e il modello predittivo creato dagli algoritmi, i cambiamenti nella formazione dell’identità e la crescente importanza degli algoritmi in questo contesto [...] stanno avvenendo all’ombra di narrazioni che lavorano per oscurare pratiche problematiche in termini economici, di giustizia e di identità.

In questo senso, la CDL opera un doppio spostamento: da un lato, come si è visto, essa non si concentra sulle nuove tecnologie da un punto di vista esclusivamente tecnico, ma le cala nel contesto economico e sociale in cui sono inserite; dall’altro, essa non si concentra esclusivamente sulla capacità del *singolo* di introiettare *skills* spendibili sul mondo del lavoro, quanto sulla capacità del *cittadino* di leggere il mondo (Freire, 1974) digitale. Ciò, a sua volta, ha due implicazioni. In primo luogo, la CDL considera la DL centrata sulle *skill* intrinsecamente ideologica, nella misura in cui questi programmi non si preoccupano di sviluppare “le capacità del cittadino di protestare, obiettare, unirsi e condurre altre azioni collettive” (Carmi *et al.*, 2020: 15) nei confronti delle pratiche di potere del capitalismo digitale. Da questo punto di vista, lo stesso approccio all’IA come tecnologia indipendente dalla concreta esperienza storico-sociale sarebbe ideologico: non riconoscere gli stretti legami tra tecnologie digitali e dinamiche di potere fa sì che, paradossalmente, la “percezione della *data literacy* sarà influenzata dai contesti sociali e politici in

cui i cittadini si troveranno a vivere e operare” (Ivi: 10). Non tematizzare i rapporti di potere da cui emergono le nuove tecnologie – e quelli a cui queste danno luogo – porta, insomma, a una loro naturalizzazione, dunque a una loro riproduzione, magari inconsapevole, nei modelli pedagogici sopra citati. In secondo luogo, l’attenzione sul *cittadino* e non sul professionista offre alla CDL la possibilità di non presentarsi come disciplina orientata allo sviluppo del singolo, quanto allo sviluppo della società. Essendo la CDL basata sulla critica delle attuali dinamiche di potere che attraversano la società digitale, essa si pone anche il problema di trascenderle, “con l’obiettivo di promuovere la giustizia sociale e il bene comune, comprendendo le relazioni e le asimmetrie di potere” (Špiranec *et al.*, 2019: np). Come ben sintetizzato da Carmi e colleghi (2020: 5), ciò che differenzia la CDL dalla DL sono, sotto questo aspetto, fondamentalmente “tre punti chiave: andare oltre l’individuo e sviluppare conoscenze a rete, sviluppare il pensiero critico riguardo l’ecosistema digitale e, da ultimo, sviluppare pratiche che diano le possibilità alle persone di diventare cittadini attivi”.

Per riassumere, dunque, la CDL si presenta come una disciplina che mira a rendere i discenti edotti non solo delle funzionalità tecniche delle nuove tecnologie, ma anche delle dinamiche di potere che le attraversano – concentrandosi sulla struttura dell’economia digitale e sulle sue costruzioni ideologiche –, con l’obiettivo di formare cittadini digitali attivi, ovvero in grado di contestare nella sfera pubblica siffatte pratiche di potere senza, tuttavia, scadere in critiche regressive o tecnofobe. Insomma, come hanno scritto Špiranec e colleghi (2019: np), la CDL si configura “come una prassi educativa situata, dove viene messo in campo un approccio critico ai dati attraverso la problematizzazione e la trasformazione di condizioni di vita oppressive e ingiuste”.

In questo contesto, emerge un problema: in che termini la CDL è *critica*? Effettivamente, come notano Carmi e colleghi (2020: 13) nella letteratura afferente alla *critical digital literacy* “non è chiaro il significato esatto di ‘critica’”. Il punto, espresso anche da Špiranec e colleghi (2019: np), è che alcuni autori si concentrano sulla critica dell’economia politica del digitale, altri sulla critica epistemologica della *data science*, altri ancora sull’ideologia e sull’“hype” che circonda l’IA e i *big data*. Insomma, possiamo effettivamente parlare di *una* CDL, o dovremmo dire che ne esistono in realtà molteplici? Esiste una dimensione critica *comune* a tutti questi lavori?

3. Cosa c’è di critico nella critical digital literacy?

Analizzando la letteratura scientifica sulla CDL è raro imbattersi in una definizione stringente di “critica”. Questo concetto tende infatti a essere utilizzato in maniera lasca e, soprattutto, non metodica. In sintesi: la CDL sancisce l’importanza dello spirito critico, ma non si interroga sulle *forme della critica* e non tematizza *cosa significhi* criticare. Senza chiarire questi presupposti metodologici, la CDL rischia di scontare gli stessi problemi che Dewey (2014) intravedeva nell’*educazione progressista*, la quale veniva accusata dal filosofo americano di fermarsi a una negazione astratta – per quanto condivisibile – dei metodi dell’“educazione tradizionale” senza offrire una reale alternativa pratico-teorica. Dal nostro punto di vista, tuttavia, la CDL possiede uno statuto critico ben definito, che va però esplicitato e tematizzato direttamente. Tramite questa operazione, diverrà anche più chiaro il ruolo cruciale che essa può svolgere all’interno di una teoria critica del capitalismo digitale. Ora, come hanno acutamente notato Špiranec e colleghi (2019, np) – anche

attraverso un'analisi quantitativa della letteratura – lo sfondo teorico, anche se non sempre consapevole, della CDL è costituito dalla pedagogia critica di Paulo Freire. A nostro avviso, esplicitare questa derivazione è fondamentale perché, in primo luogo, permette di comprendere non solo *che genere di critica* sia in gioco, ma anche perché siffatta critica è utile in un contesto digitale. Inoltre, dato che la riflessione di Freire ben si integra con i risultati e i programmi di ricerca della teoria critica contemporanea, esplicitare lo sfondo freiriano della CDL ci permetterà di integrarla all'interno di una teoria critica del capitalismo digitale.

Ne *L'educazione come pratica di libertà*, Freire scrive: “quando l'uomo coglie i dati oggettivi della realtà o i rapporti che legano un fatto a un altro, il suo agire diventa naturalmente critico, cioè riflessivo” (Freire, 2024: 47). L'agire critico, dunque, si configura come un agire capace di cogliere i dati oggettivi della realtà e i loro rapporti, smascherandone la natura ideologica e superando “la falsa coscienza del mondo” (Freire, 2019: 96). Critica, qui, non significa mera denuncia, ma – come specifica Freire – agire *riflessivo*, in senso rigorosamente hegeliano (Hegel, 2006): la riflessione va infatti intesa come quel movimento del pensiero che *toglie* (*Aufheben*) l'immediatezza e la *fatticità* della situazione, svelando il suo essere *prodotta* da concrete pratiche umane e il suo essere attraversata da linee di forza e di potere (Marcuse, 2020). Per mettere in campo un agire critico, tuttavia, gli esseri umani non possono agire *indipendentemente* dalla realtà, secondo qualche principio trascendente. Al contrario, scrive Freire, è necessario *integrarsi* nel e col mondo. Si tratta di “entrare in accordo con la realtà e al contempo di trasformarla” (Freire, 2024: 49). L'agire critico-riflessivo non può prescindere dalla consapevolezza del dramma di *questo mondo*. Non può ignorarlo e pensarlo a partire da uno sguardo da nessun luogo. Al contrario, solo l'integrazione col mondo, ovvero la conoscenza delle strutture

che lo governano, permette al soggetto di non diventare mero “zimbello dei cambiamenti” (Ivi: 52). La pedagogia critica di Freire, dunque, non si appella a presunti valori universali. Al contrario, essa muove dall’integrazione dei discenti col e nel mondo, ma mira a far sì che essi sviluppino quella che Freire chiama “transitività critica”, la quale “si caratterizza per il *modo* di interpretare i problemi, per la sostituzione delle spiegazioni magiche con principi di causa” (Ivi: 67). Il soggetto, insomma, deve essere pronto ad accogliere dentro di sé il mondo – dunque *deve conoscerlo* – ma deve anche essere in grado di trattarlo riflessivamente, contestando e mettendo in discussione ogni pregiudizio e verificando ogni informazione che riceve, non accontentandosi di spiegazioni “magiche”, ma cercando sempre di ricostruire le catene causali che hanno generato un determinato fatto sociale. In questo senso, secondo Freire la “criticità” non è riducibile a un mero spirito di opposizione. Al contrario, essa “significa il suo inserimento, la sua integrazione, la sua maniera oggettiva di vedere la realtà” (Ibidem), in virtù delle quali i soggetti tolgono l’immediatezza della realtà, la mettono in discussione e, dunque, sviluppano la possibilità di trasformarla. È importante sottolineare come la dimensione critica della pedagogia freiriana non abbia un contenuto *determinato*, il quale dovrà emergere innanzitutto dalla prassi dialogica che questa punta a sviluppare. Piuttosto, essa mira innanzitutto a produrre nei soggetti “un tentativo costante di cambiare atteggiamento” (Ivi: 96) nei confronti della realtà, che – lungi dall’essere intesa come qualcosa di dato una volta per tutte – deve essere sempre intesa come “problema” (Freire, 2019: 94).

Come notano Bevilacqua e Casarin (2024: 10), “l’educazione come pratica di libertà emerge dunque non come un metodo, ma come una disposizione di attenzione e allerta alle dinamiche di dominio a cui siamo soggetti”. Non si tratta di promuovere un’immagine di vita buona, quanto di sottolineare “l’urgenza di

un'attitudine critica, la necessità di approcci problematizzanti e il valore inestimabile di rinnovate pratiche dialogiche" (Kohan, 2023: 15), fondate sulla capacità dell'educazione di attivare la riflessività dei soggetti. Ora, su queste basi teorico-filosofiche, Freire oppone educazione problematizzante ed educazione depositaria. Se la prima è basata sull'integrazione col mondo, sulla riflessività e la transattività critica e ha come obiettivo principale quello di aumentare "il campo della percezione" (Freire, 2019: 92) dei soggetti, facendo sì che "ciò che prima esisteva oggettivamente [...] assume il carattere di problema" (Ivi: 92), la seconda, invece, "insiste nel mantenere occulte certe ragioni" (Ibidem). La concezione depositaria dell'educazione, secondo Freire, vela la storicità delle cose, il loro essere prodotte e il loro essere figlie di rapporti di potere: essa parla "della realtà come qualcosa di fermo, statico, suddiviso e disciplinato" (Ivi: 77) utilizzando un tono al contempo narrativo – in virtù del quale l'educazione diventa l'atto di depositare" (Ivi: 78) – e astratto, nella misura in cui essa sottopone ai discenti "contenuti che sono dei veri e propri ritagli della realtà, sconnessi rispetto all'insieme da cui hanno origine e in cui troverebbero significato" (Ivi: 77). Insomma, se l'educazione problematizzante va intesa come un "atto di conoscenza" (Ivi: 88) che "stimola la riflessione e l'azione autentica dell'uomo sulla realtà" (Ivi: 93), l'educazione depositaria – nella misura in cui "maschera la realtà con dei miti" (Ivi: 92) e opera "una specie di anestesia" (Ivi: 90) – mantiene i discenti in una condizione di *adattamento* rispetto al mondo e, dunque, "pretende, nell'ambito della sua definizione ideologica, di addottrinarli affinché si adattino al mondo dell'oppressione" (Ivi: 86). L'educazione depositaria si configura allora come estensione delle pratiche di dominio, nella misura in cui il reale viene affrontato astrattamente e ideologicamente. La realtà e le dinamiche di potere che l'attraversano, lungi dall'essere intese come problemi, agiscono qui

come criteri di verità e giustizia non indagati riflessivamente. La conseguenza, conclude Freire, è che siffatto genere di educazione “soddisfa gli interessi degli oppressori” (Ivi: 81), nella misura in cui “l’unico margine d’azione che si offre agli educandi è ricevere i depositi, conservarli e metterli in archivio” (Ibidem). Ora, alla luce di questa distinzione freiriana, la critica della CDL alla DL appare decisamente più fondata. Per come si presenta, infatti, quest’ultima esibisce in maniera chiara i tratti dell’educazione depositaria. Le obiezioni della CDL nei confronti della *digital literacy*, infatti, hanno per oggetto esattamente la sua dimensione *adattiva*: come si è visto, la DL non mira a formare cittadini in grado di comprendere attivamente e criticamente le trasformazioni tecnologiche, politiche ed economiche portate dalla rivoluzione digitale. Al contrario, essa mira a formare *individui adatti al mondo del lavoro* che, padroneggiando l’uso delle nuove tecnologie, sappiano utilizzarle efficacemente per aumentare la loro competitività e la loro impiegabilità. Peraltro, il fatto stesso che la DL si pensi, innanzitutto, come *literacy*, dunque come alfabetizzazione, mostra come l’acquisizione di *skills* digitali sia considerato prioritario rispetto a una conoscenza critico-riflessiva dell’ambiente sociotecnico in cui i cittadini si trovano a operare. Sotto questo punto di vista, la DL si configura – per quanto a volte inconsapevolmente – come un approccio ideologico, giacché l’immediatezza della realtà digitale non è praticamente *mai* messa in questione, né ai discenti viene offerta la possibilità di esperire quella che Freire chiama *integrazione* col e nel mondo, ovvero un *cambio d’atteggiamento* che gli permetta di criticare l’attuale stato di cose proprio in virtù della sua comprensione. Al contrario, l’attuale ecosistema digitale viene considerato come un che di dato una volta per tutte, scevro da contraddizioni e rapporti di potere, e, dunque, come un futuro prestabilito a cui non resta che adattarsi, attraverso la deposizione – nelle teste dei discenti – di competenze

strettamente tecniche che mirano a renderli *fluente* in *computer science*. Peraltro, anche nei casi in cui la DL fa riferimento a questioni non strettamente tecniche – come la “storia dell’AI” prevista dal framework di Su e Zhong (2022) o le “five big ideas” tematizzate da Touretzky e colleghi (2019) –, essa tende ad assumere un atteggiamento narrativo, se non addirittura agiografico per quanto riguarda la gloriosa storia delle nuove tecnologie e dei suoi eroi, che Freire ben descrive all’inizio del capitolo secondo de *La Pedagogia degli oppressi* e che si configura come “narrazione di contenuti, che per ciò stesso tendono a fossilizzarsi [...]. La parola, in queste dissertazioni, si svuota della dimensione concreta che dovrebbe avere, o si trasforma in bla-bla-bla, in verbosità alienata e alienante” (Freire, 2019: 76). Insomma, intendere la CDL come una forma di educazione problematizzante ci permette di criticare la DL anche in quanto forma di educazione depositaria, dunque in quanto forma di alfabetizzazione che, per quanto necessaria, nel suo essere astratta e poco interessata a ciò che avviene *dietro* gli artefatti tecnologici, si configura in fin dei conti non solo come una pratica ideologica, ma anche come una pratica che non permette ai discenti di ampliare il raggio della loro esperienza e di sviluppare quelle capacità critico-riflessive che sarebbero necessarie per porre la questione di una governance effettivamente emancipativa della rivoluzione digitale. Sotto questo aspetto, l’impegno dell’educazione problematizzante a superare l’immediatezza della realtà e a ricostruire i processi genetici della situazione in cui siamo gettati può senza dubbio agire, dal punto di vista della CDL, come “unificatore concettuale”, nella misura in cui quest’ultima – anche se non sempre esplicitamente – si pensa esattamente come una prassi critico-pedagogica che mira ad andare *oltre* il mero fatto tecnico, tematizzando il rapporto tra *media* e società, tra tecnologie digitali e loro rappresentazioni ideologiche e tra algoritmi, IA e soggettività umane. Infine, è

fondamentale sottolineare come la conoscenza offerta dall'educazione problematizzante – che mira svelare quelle pratiche di dominio che sono nascoste dietro le ideologie dell'oggettività, della velocità e della *smartness* algoritmica – non possa essere intesa come mera presa d'atto di dati di fatto. Al contrario, siffatta pratica deve essere pensata come la volontà di sviluppare una *conoscenza riflessiva orientata all'emancipazione*. Dunque, la chiarificazione dello statuto critico della CDL che abbiamo svolto ci porta non solo a comprendere meglio la sua potenza critica nei confronti della DL, ma anche a capire *per quale motivo* essa può svolgere un ruolo centrale all'interno di una più ampia teoria critica del capitalismo digitale.

4. Il ruolo della critical digital literacy in una teoria critica del capitalismo digitale

Nel primo paragrafo, si è visto come il capitalismo digitale abbia una fortissima dimensione ideologico-cognitiva, in virtù della quale esso tende a produrre soggetti caratterizzati da “deficit strutturali di riflessività” (Celikates, 2018: 175) – ovvero non abituati a esercitare un pensiero critico nei confronti delle strutture di senso e di potere in cui sono immersi – con l'obiettivo di fluidificare il processo economico e moltiplicare così i profitti (Srnicek, 201; Langley & Leyshon, 2016; Barry, 2019; Zuboff, 2019). Evidentemente, la *digital literacy*, intesa nel senso freiriano di educazione depositaria, concorre, insieme ai dispositivi messi in campo dal capitalismo digitale (mito dell'oggettività algoritmica, *digital nudging*, accelerazione della vita quotidiana), a mantenere i soggetti in questo stato, giacché – come si è visto – le tecnologie digitali vengono prese in considerazione in maniera astratta (ovvero slegandole dal contesto economico e sociale in cui agiscono) e

ideologica, ovvero concentrandosi soprattutto sulle *skills* che i soggetti devono acquisire per essere competitivi nel mondo del lavoro. La CDL propone evidentemente un modello alternativo, in cui l'approccio nei confronti degli artefatti tecnologici non è volto alla mera capacità strumentale, ma punta anche, se non soprattutto, a toglierne l'apparente immediatezza, considerando il discente non come singolo professionista ma come cittadino, e portandolo a interrogarsi sulle concrete forme di potere e dominio che si celano dietro alla presunta neutralità, oggettività e onnipotenza delle nuove tecnologie. Ovviamente, la CDL insiste moltissimo anche sui fattori strettamente tecnici ed epistemologici, ma ciò che la differenzia dalla DL è che il suo obiettivo non è solo quello di formare soggetti in grado programmare meglio un'IA o di scrivere meglio un algoritmo. Al contrario, avere competenze tecniche è cruciale per comprendere quando le narrazioni sugli algoritmi o sull'AI trascendono le loro effettive capacità, e questo è decisivo per riuscire a scovare i *clivages* ideologico-cognitivi su cui il capitalismo digitale fa leva per realizzare i propri obiettivi economici. Insomma, sapere che gli algoritmi non sono neutri perché riproducono discriminazioni *già presenti* nella società in cui operano permette non solo di mitigare questi *bias*, ma anche di comprendere quando un'iniziativa aziendale si configura in realtà come una forma di *ethic* o *pinkwashing*, pratiche sempre più diffuse nelle grandi aziende digitali e che raramente vengono percepite come tali dagli utenti (Zuboff, 2019; Numerico, 2021). Allo stesso modo, essere a conoscenza del problema dell'implementazione o di quello dell'approssimazione (cfr. *supra*, §1), permette ai soggetti di cogliere la dimensione ideologica di narrazioni come quella dell'oggettività algoritmica, mentre conoscere il modo in cui una piattaforma organizza il suo *design* permette di mitigare la forza dei *digital nudge*, i quali operano meglio nel silenzio e nell'inconsapevolezza (Yeung, 2017).

Insomma, la CDL è cruciale nello sviluppo di una teoria critica del capitalismo digitale perché permette ai soggetti di mettere in campo pratiche riflessive, dotandoli di una conoscenza che, pur tecnicamente accurata, non si esaurisce nella mera conoscenza del fatto tecnico. Per recuperare una categoria habermasiana, la CDL si configura infatti come una conoscenza orientata all’emancipazione, nella misura in cui essa contesta proprio una “condizione della comunicazione *sistematicamente distorta* e [...] una repressione apparentemente legittimata” (Habermas, 1990: 321). Nella misura in cui si danno pratiche ideologiche volte a distorcere la comunicazione intersoggettiva e a legittimare la repressione delle strutture di senso del mondo di vita, la “forza emancipativa della riflessione” (Ivi: 194) – che, “trasformatasi in critica, si è del tutto liberata dall’accecamento ideologico” (Ivi: 59) – fa coincidere la conoscenza del dato, intesa innanzitutto come *toglimento* della sua immediatezza, “con l’interesse all’emancipazione” (Ivi: 194). Non si tratta, qui, di proporre contenuti specifici ai discenti, né di pretendere – in maniera paternalista o elitista, come troppe volte è capitato alla teoria critica – di dettare la linea dell’agire sulla base di una presunta conoscenza totalizzante del decorso storico. Al contrario, la CDL avrebbe il compito, coerente con le indicazioni freiriane, di instillare nei soggetti un atteggiamento riflessivo e al contempo non definitivo, sempre aperto al dialogo e alla problematizzazione, configurandosi – secondo una felice metafora adorniana – come un pensiero, ma si potrebbe dire un modo di vivere, caratterizzato dall’“insistenza” (Adorno, 2007: 194), ovvero da quella *Haltung* critica che è oramai “passata nella forma della resistenza, dell’opposizione contro il mondo, della pazienza di cui si dà prova quando non si accondiscende all’ideologia” (Ivi: 192). L’obiettivo della CDL, da questo punto di vista, non è in alcuna maniera “tecnofobo” o “catastrofista”. Al contrario, essa avrebbe il compito di ricordare ai soggetti che il punto di vista tecnico-

scientifico, in tutta la sua *potenza* (che non può essere negata astrattamente), “deve essere, poi, criticato nei suoi stessi limiti, ‘scoperto’ nel suo non-detto e nei suoi presupposti, collocato realisticamente nelle sue relazioni con la situazione storico-politica” (Cacciari, 2020: 44). Il punto è far sì che lo sviluppo tecnologico non proceda isolatamente, come una forma invariante e necessaria che, assomigliando al “modello dell’automa” (Habermas, 1990: 59), nasconde in realtà concrete pratiche di dominio. Al contrario, ciò che una teoria critica del capitalismo digitale dovrebbe fare è riuscire, in qualche modo, a far sì che siffatto avanzamento tecnico sia mediato “da gradi della riflessione attraverso i quali la dogmatica di forme superate di ideologie e di dominio viene dissolta, la pressione del quadro istituzionale sublimata e l’agire comunicativo liberato *come* comunicativo” (Ibidem). L’obiettivo, lo si è detto più volte, è garantire la possibilità di un’appropriazione critico-riflessiva delle nuove tecnologie, con il fine di rendere possibile un *discorso* critico orientato all’emancipazione che permetta ai soggetti ordinari di porre il tema di una governance emancipativa, e non sistematicamente distorta e mediata dal dominio, dei mezzi tecnologici, senza tuttavia scadere in posture tecnofobe o neoluddiste. In questo senso, la CDL avrebbe il ruolo, all’interno di una teoria critica del capitalismo digitale, di operare come prassi al contempo di rinforzo e di tutela della riflessività dei soggetti, evitando così che si crei una frattura teorico-pratica tra il “critico” – riflessivo e consapevole – e gli agenti ordinari – inconsapevoli e persi in un intrascendibile *digital habitus* (Celikates, 2018; Jaeggi, 2021). Per quanto questo obiettivo possa apparire limitato e di certo non raggiungibile nel breve periodo, esso è in realtà cruciale e ne va della possibilità stessa della critica che – qualsiasi forma assuma – non può fare a meno di fondarsi su pratiche situate di riflessività, ovvero sulla capacità *degli agenti ordinari* di cogliere l’“inaccettabilità riflessiva” (Celikates, 2018:

158) delle forme di dominio che caratterizzano la loro realtà sociale. E, in un contesto come quello del capitalismo digitale, una tale operazione appare più necessaria che mai, nella misura in cui esso agisce continuamente proprio per chiudere lo spazio di riflessività dei soggetti, *conditio sine qua non* per la realizzazione dei suoi obiettivi commerciali. Solo sulla base di una riattivazione della riflessività degli agenti ordinari è poi possibile pensare forme utopiche di riorganizzazione sociale che, negando in maniera determinata le forme di dominio senza rinunciare alla potenza della tecnica, permettano ai soggetti di “mobilitare le risorse dell’immaginazione in direzione dell’emancipazione” (Mordacci, 2023:146). In questo senso, la CDL – intesa come conoscenza riflessiva orientata all’emancipazione – è anche catalizzatrice del pensiero utopico, nella misura in cui è la presa in considerazione delle contraddizioni sociali a generare “l’immaginazione dell’alternativa ed è da questa che si trae lo spunto e la motivazione per il cambiamento” (Ibidem). Soprattutto, l’umiltà epistemica della CDL – che non specifica mai cosa debba essere fatto, sulla base di una presunta totalizzazione da parte del critico del mondo sociale o del processo storico – apre alla possibilità di una *moltitudine* di utopie, tutte egualmente critiche (almeno nella genesi e nei presupposti trascendentali di giustizia, rispetto e uguaglianza) che – proprio per il loro essere *frammentarie* (Ibidem) – possono, e devono, entrare *in dialogo*. Solo nel *dialegesthai* tra utopie critiche, infatti, è possibile ripensare – nel nostro mondo mediatizzato – non solo lo sviluppo di una conoscenza orientata all’emancipazione, ma anche la ricostruzione di una sfera pubblica effettivamente capace di agire *politicamente*, ovvero di innestare nel dibattito pubblico, specie se distorto dal dominio, tematiche eticamente pregnanti o da esso sistematicamente escluse (Fraser, 1990). Sotto questo aspetto – seppur nel lungo periodo – la CDL può certamente giocare un ruolo

cruciale, oltre che pienamente allineato alle necessità che l'attuale congiuntura impone al pensiero critico.

Bibliografia

Adorno, T. W. (2007). *Terminologia Filosofica*. Torino: Einaudi.

Angius, N. (2023). *The philosophy of computer science*, Stanford Encyclopedia of Philosophy, 19/01/2023: <https://plato.stanford.edu/ENTRIES/computer-science/#Bib>.

Barry, L. (2019), The rationality of the digital governmentality. *Journal for cultural research*, 1, 365-380.

Beer, D. (2019). *The Data Gaze*. London: SAGE.

Ben-David, S. *et al.* (2019). Learnability Can Be Undecidable, *Nature Machine Intelligence*, 1, 44-48.

Berns, T., Rouvroy, A. (2013). Algorithmic governmentality and prospects of emancipation. *Réseaux*, 1, 163-196.

Bevilacqua, S. & Casarin, P. (2024). *Premessa*. In P. Freire, *Educazione come pratica di libertà*. Milano; Udine: Mimesis.

Bourdieu, P. (1990). *The logic of practice*. Stanford: Stanford University Press.

Cacciari, M. (2020). *Il lavoro dello Spirito*. Milano: Adelphi.

Carmi, E. *et al.* (2020). Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation. *Internet Policy Review*, 9(2), 1-22.

Carrington, V. (2018). The changing landscape of literacies: Big data and algorithms. *Digital Culture and Education*, 10(1), 67-76.

Celikates, R. (2018). *Critique as social practice*. London: Rowman & Littlefield.

Cheney-Lippold, J. (2017). *We are data*. New York: New York University Press.

Cohn, J. (2019). *The burden of choice*. New Brunswick: Rutgers University Press.

Csikszentmihalyi, M. (2008). *Flow: The Psychology of Optimal Experience*. New York: Harper & Row.

Darmody, A., Zwick, D. (2020). Manipulate to empower: Hyper-relevance and the contradictions of marketing in the age of surveillance capitalism. *Big Data & Society*, 1, 1-12.

Delanty, G., Harris, N. (2021). Critical theory and the question of technology: The Frankfurt School revisited. *Thesis Eleven*, 1, 88-108.

Dewey, J. (2014). *Esperienza ed educazione*. Milano: Cortina.

Finn, E. (2017). *What algorithms Want*. Cambridge: MIT Press.

Fisher, E. (2022). *Algorithms and Subjectivity*. New York: Routledge.

Fraser, N. (1990). Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy. *Social Text*, 25/26, 56-80.

Freire, P. (2019). *Pedagogia degli oppressi*. Torino: Edizioni Gruppo Abele.

Freire, P. (2024). *Educazione come pratica di libertà*. Torino: Einaudi.

Fuchs, C. (2021). *Marxist Humanism and Communication Theory*. New York: Routledge.

Fumagalli, A. (2007). *Bioeconomia e capitalismo cognitivo*. Verona: Ombre Corte.

Gillespie, T. (2016). Algorithm. In Peters, E. (ed), *Digital Keywords*. Princeton: Princeton University Press. 18-30.

Habermas, J. (1990). *Conoscenza e Interesse*. Roma/Bari: Laterza.

Hegel, G.W.F. (2006). *Scienza della Logica*, Roma/Bari: Laterza.

Hristova, S. (2023). Habits and habitus in digital culture. In Lindgren, S. (ed.). *Handbook of critical studies on artificial intelligence*. Cheltenham: Elgar. 108-117.

Illis, K., Petit, M., Jarret, K. (2013). *Google and the culture of search*. New York: Routledge.

Jaeggi, R. (2021), *Critica delle forme di vita*. Milano/Udine: Mimesis.

Kandlhofer, M. *et al.* Artificial intelligence and computer science in education: From kindergarten to university. In *2016 IEEE frontiers in education conference (FIE)*. <https://ieeexplore.ieee.org/abstract/document/7757570>.

Karakayali, *et al.* (2018). Recommendation Systems as Technologies of the Self. *Theory, Culture & Society*, 2, 3-24.

Kohan, W. (2023). *Paulo Freire, più di sempre. Una biografia filosofica*. Milano/Udine: Mimesis.

Langley, P., & Leyshon, A. (2016). Platform Capitalism: The Intermediation and Capitalization of Digital Economic Circulation. *Finance and Society*, 1, 11-31.

Lindgren, S. (2023). *Critical theory of AI*. New York: Wiley.

Lupton, D. (2016). *The Quantified Self: A Sociology of Self-Tracking*. Cambridge: Polity Press.

Marcuse, H. (2020). *Ragione e Rivoluzione. Hegel e il sorgere della teoria sociale*. Bologna: Il Mulino.

Mordacci, R. (2023). *Critica e Utopia*. Roma: Castelvecchi.

Ng, D.T.K., *et al.* (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 1-11. <https://doi.org/10.1016/j.caeai.2021.100041>.

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

Numerico, T. (2021). *Big Data e Algoritmi. Prospettive Critiche*. Roma: Carocci.

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

Pfeiffer, S. (2022). *Digital Capitalism and Distributive Forces*. Bielefeld: Transcript Verlag.

Pierazzuoli, G. (2022). *Il soggetto collaborativo. Per una critica del capitalismo digitale*. Verona: OmbreCorte.

Possati, L.M. (2021). *The Algorithmic Unconscious*. New York: Routledge.

Romele, A. (2023). *Digital Habitus*. New York: Routledge.

Rouvroy, A. (2016). Algorithmic governmentality: Radicalisation and immune strategy of capitalism and neoliberalism? *La deleuziana*, 3(1), 30-36.

Scaltsas, T. (2017), "Valuative Intelligence". *Medienimpulse*, 4, 1-11.

Špiranec, S. *et al.* (2019). "Searching for critical dimensions in data literacy", in *Proceedings of the Tenth International Conference on Conceptions of Library and Information Science, Ljubljana, Slovenia*,

June 16-19, 2019. <https://informationr.net/ir/24-4/colis/colis1922.html>.

Srnicek, N. (2016). *Platform Capitalism*. Cambridge: Polity Press.

Stiegler, B. (2017). *Automatic Society: The Future of Work*. Cambridge: Polity Press.

Touretzky, D., et al. (2019). Envisioning AI for K-12: What should every child know about AI?. *Proceedings of the AAAI conference on artificial intelligence*, 1. 9795-9799.

Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, 12(2). 197-208.

Wajcman, J. (2014). *Pressed for Time: The Acceleration of Life in Digital Capitalism*. Chicago: University of Chicago Press.

Yeung, K. (2017). 'Hypernudge': Big Data as a Mode of Regulation by Design. *Information, Communication & Society*, 20(1). 118-136.

Zellini, P. (2010). *Numero e Logos*. Milano: Adelphi.

Zuboff, S. (2019). *The Age of Surveillance Capitalism*. New York: PublicAffairs.

4. TRASFORMAZIONI DIGITALI NEL WELFARE: INTELLIGENZA ARTIFICIALE E SERVIZIO SOCIALE

di Roberto Veraldi*, Chiara Fasciani**

Abstract: Negli ultimi anni, l'intelligenza artificiale è emersa come una delle tecnologie più trasformative del nostro tempo, portando a una riorganizzazione delle relazioni sociali e del sistema di Welfare. Ciò rende necessario un ripensamento delle politiche sociali e dei servizi destinati alle persone con disabilità e altre fragilità. Questo studio propone di analizzare l'impatto dell'IA sui processi decisionali e sul lavoro degli operatori sociali, valutando sia i benefici che i potenziali rischi.

Parole chiave: Intelligenza Artificiale, Welfare, Servizi Sociali, Politiche Sociali.

Abstract: In recent years, artificial intelligence has emerged as one of the most transformative technologies of our time, leading to a reorganization of social relationships and a restructuring of the Welfare System. This makes it necessary to rethink social policies and services for people with disabilities and other fragilities. This study aims to analyze the impact of AI on decision-making processes and on the actions of social workers, assessing both the benefits and the potential risks.

Keywords: Artificial Intelligence, Welfare, Social Services, Social policies.

* Associate professor presso il Dipartimento di Scienze Filosofiche, Pedagogiche ed Economico-Quantitative, Università degli Studi "Gabriele d'Annunzio" di Chieti-Pescara, roberto.veraldi@unich.it.

** Ricercatrice a contratto presso il Dipartimento di Scienze Filosofiche, Pedagogiche ed Economico-Quantitative, Università degli Studi "Gabriele d'Annunzio" di Chieti-Pescara, chiara.fasciani@collaboratori.unich.it.

Accettato Dicembre 2024 - Pubblicato Aprile 2025.

Introduzione

Negli ultimi anni, l'intelligenza artificiale (IA) è emersa come una delle tecnologie più influenti e trasformative del nostro tempo, capace di rivoluzionare vari settori della società e del sistema di welfare, ridefinendo i processi, migliorando l'efficienza e aprendo nuove possibilità (Kissinger, 2023). Questo tipo di applicazione delle tecnologie segna una trasformazione significativa nel contesto della società post-moderna, caratterizzata di per sé da una crescente complessità e pluralismo, con una discontinuità rispetto ai modelli tradizionali di organizzazione sociale (Ardigò, 1988). In questo scenario, l'intelligenza artificiale non solo offre opportunità per l'innovazione nel welfare, ma solleva anche interrogativi cruciali sulla giustizia sociale, l'accesso alle risorse e l'autonomia degli individui. Il potenziale dirompente delle nuove tecnologie è in grado di provocare cambiamenti radicali, addirittura nell'ambito della fede e della spiritualità, portando anche a parlare di religione digitale, pur con le dovute distorsioni del caso, come sottolinea anche Roberto Cipriani (2020). Dal nostro punto di vista, questo avanzamento porta con sé una serie di questioni etiche, politiche e sociali che richiedono attenta considerazione e che sono state anche al centro della discussione del Forum G7 del 14 giugno 2024, dove uno degli interventi più significativi è stato quello di Papa Francesco che ha sottolineato l'importanza di un approccio etico nello sviluppo e nell'uso dell'automazione avanzata, richiamando il concetto di "Algor-etica" (Benanti, 2018): un ponte che integra i principi di dignità umana, solidarietà e giustizia nelle tecnologie digitali attraverso un dialogo interdisciplinare. All'interno di questo lavoro, si prevede di analizzare criticamente l'impatto dell'IA sui processi decisionali e sui servizi in ambito assistenziale, valutando sia i benefici che i potenziali rischi. Alston parla di «welfare state digitale» per riferirsi ai tentativi di rendere

più efficienti e mirati i servizi prestati attraverso le nuove tecnologie e l'intelligenza artificiale, soprattutto in riferimento alla gestione delle fragilità (Parra-Saiani, 2020; Alston, 2019). Per le persone con disabilità, le tecnologie offrono strumenti di supporto avanzati che aumentano l'autonomia e la qualità della vita attraverso dispositivi intelligenti e soluzioni personalizzate. Nell'assistenza agli anziani, l'IA promuove l'invecchiamento attivo e indipendente: robot di compagnia e sistemi di monitoraggio remoto sono impiegati per migliorare sicurezza e benessere, riducendo al contempo il carico sui caregiver. Le soluzioni digitali, incluse piattaforme online e applicazioni mobili, trasformano l'accesso ai servizi sociali, ottimizzano l'efficienza amministrativa e consentono di analizzare grandi quantità di dati per prevedere trend sociali, individuare gruppi vulnerabili e proporre interventi mirati, facilitando una gestione proattiva del welfare. Tuttavia, l'integrazione della digitalizzazione e dell'IA nei sistemi di welfare non è priva di sfide: vari studiosi e attivisti avvertono che tali iniziative potrebbero aggravare le disuguaglianze e la discriminazione, piuttosto che ridurle, minando ulteriormente i diritti delle popolazioni già emarginate. (Parra-Saiani, 2020; Alston, 2019). Partendo da queste brevi premesse, emerge chiaramente la necessità di identificare soluzioni e strategie per massimizzare i benefici e minimizzare l'impatto sociale negativo delle nuove tecnologie nel welfare.

1. Intelligenza artificiale e prime definizioni

L'espressione "Intelligenza Artificiale" si riferisce a sistemi o macchine in grado di imitare l'intelligenza umana durante l'esecuzione di specifici compiti per i quali sono programmati e che possono migliorarsi in itinere sulla base della raccolta e della

rielaborazione di dati e informazioni (Russel & Norvig, 2005). L'IA è stata definita dalla Commissione Europea come: «*Systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals*» (European Commission, 2018). Questa definizione mette in evidenza due aspetti fondamentali: la capacità dei sistemi di analizzare l'ambiente e la loro autonomia nell'agire per raggiungere obiettivi specifici. L'enfasi sul "comportamento intelligente" implica che l'IA non si limita all'elaborazione di dati preesistenti, ma include l'adattamento e la risposta dinamica a situazioni inedite. Inoltre, la "certa autonomia" dei sistemi, sottolinea l'importanza della decisione automatica e della capacità di operare senza supervisione continua. Nel 2019, il Gruppo di Esperti sull'Intelligenza Artificiale presso la Commissione Europea fornisce un'ulteriore definizione, considerandole come sistemi che, dato un obiettivo complesso, agiscono attraverso l'acquisizione e l'interpretazione di dati, ragionando su di essi e decidendo le migliori azioni da intraprendere per raggiungere tale obiettivo (High-Level Expert Group on AI, 2019a). Ad oggi, l'IA non pensa (e questo dovrebbe anche farci riflettere sull'utilizzo del termine "intelligenza" per definire questo strumento), ma rielabora informazioni e costruisce percorsi logici grazie alla sua capacità di attingere a enormi bacini di dati e selezionare quelli utili e congruenti con le richieste. Di conseguenza, negli ultimi anni si è assistito a un rinnovato interesse verso il settore dell'automazione intelligente e dei processi computazionali avanzati, proprio grazie alla crescita vertiginosa della mole di dati disponibili e agli sviluppi tecnologici correlati, che hanno supportato e potenziato le capacità umane, principalmente in termini di efficacia ed efficienza.

2. Aspetti generali legati all'impatto sociale dell'IA

L'adozione diffusa dell'intelligenza artificiale ha avuto un impatto profondo e diversificato sulla società e sulla cultura che si estende a numerosi settori, trasformandoli in modi significativi, ma sollevando anche nuove questioni sociali ed economiche. In particolare, in ambito sanitario gli algoritmi di apprendimento automatico consentono di individuare malattie precocemente e personalizzare le cure, inoltre supportano la telemedicina, l'assistenza domiciliare e la gestione delle malattie croniche (Panesar, 2019). Oltre ai benefici economici e funzionali, l'IA impatta anche le dinamiche sociali e lavorative, modificando le interazioni quotidiane e facilitando il lavoro remoto attraverso strumenti digitali come videoconferenze e piattaforme collaborative. L'emergere recente dell'intelligenza artificiale generativa, però, ha sollevato interrogativi su quanto siamo vicini a una rapida accelerazione nell'automazione dei compiti e alla riduzione dei costi del lavoro. Ciò si traduce in un cambio di prospettiva circa il panorama lavorativo, contribuendo al tempo stesso ad un aumento della disoccupazione tecnologica e causando la perdita di posti di lavoro in settori caratterizzati da compiti ripetitivi e manuali (Hatzius, 2023). Inoltre, le trasformazioni del mercato del lavoro, guidate da innovazioni tecnologiche come l'affermazione dell'intelligenza artificiale generativa, sono amplificate da interruzioni economiche e geopolitiche e da crescenti pressioni sociali (Zahidi, 2023). La frontiera tra uomo e macchina si è spostata e secondo il rapporto "Future of Jobs Report 2023", si stima che a livello globale il 66% delle mansioni lavorative sia attualmente svolto da esseri umani, mentre il restante 34% è attribuito a robot; inoltre, si prevede una crescita della percentuale di lavori eseguiti da robot negli anni a venire e un aumento della disoccupazione umana (World Economic Forum, 2023). In particolare, l'evoluzione

tecnologica sta generando nuovi ruoli e professioni focalizzati sulla gestione, sviluppo e manutenzione dei sistemi di IA, che richiedono competenze avanzate in informatica, ingegneria e analisi dei dati. Questo cambiamento richiede interventi mirati nel welfare e nelle politiche attive del lavoro, volti a sostenere la transizione professionale e a garantire l'adeguamento delle competenze, favorendo così l'occupabilità in un mercato sempre più orientato alle tecnologie avanzate. Si pensi, ad esempio, a programmi di riqualificazione, politiche di supporto per i lavoratori disoccupati e a percorsi di formazione continua per l'intera forza lavoro.

3. Welfare, social work e digitalizzazione: una panoramica

La digitalizzazione ha portato, in generale, a una rimodulazione delle relazioni sociali e ad una riorganizzazione profonda dei sistemi di welfare e dei servizi sociali, richiedendo una riprogettazione delle politiche sociali e dei modelli di assistenza. L'IA si sta affermando in vari settori e l'ampia gamma delle sue applicazioni può risultare dispersiva. Per questo motivo, si propone una prima categorizzazione strutturata (Tab. 1) per comprendere le principali aree di intervento.

Tabella 1: Tabella riassuntiva delle applicazioni dell'IA nel welfare

Ambito	Applicazione IA	Esempi
Programmazione, Progettazione e Gestione dei servizi	Analisi predittiva, Progettazione e Allocazione risorse	Programma “One View” (Londra)
Automazione ed efficienza dei Servizi	Chatbot, gestione digitale	AI nei servizi pubblici, Welfare Aziendale digitalizzato
Supporto alla salute e al benessere bio-psico-sociale	Monitoraggio remoto, telemedicina	Care Technology, Smart home technology
Inclusione sociale di persone fragili	Tecnologie assistive, robot sociali	Sostegno e inclusione di anziani e disabili

Fonte: Elaborazione a cura degli autori

3.1 IA per il miglioramento della programmazione, progettazione e gestione dei servizi sociali

L’IA consente di ottimizzare la gestione delle risorse pubbliche e private destinate al welfare, migliorando la capacità di risposta ai bisogni emergenti e trasformando la programmazione sociale in una realtà “smart”. Nel settore pubblico, può identificare modelli e tendenze all’interno dei dati, permettendo agli operatori dei servizi sociali di anticipare le esigenze future e di allocare le risorse in modo più efficace. Le tecniche di *machine learning* consentono di elaborare grandi volumi di dati strutturati e non strutturati, prevedere tendenze sociali, identificare gruppi vulnerabili e suggerire interventi mirati. Questa capacità predittiva consente una gestione proattiva del welfare, migliorando i tempi di risposta e l’efficacia degli interventi: attraverso la triangolazione dei dati è possibile sviluppare previsioni, analizzare gli scenari e determinare le possibilità di successo (McFarland, 2024). L’analisi predittiva potrebbe rivelarsi come una delle applicazioni più potenti

dell'IA perché, di fatto, consente di anticipare gli eventi futuri e le esigenze della popolazione, grazie alla raccolta e all'elaborazione di dati storici e attuali. Un esempio significativo è rappresentato dal “*London Borough of Barking and Dagenham*”, uno dei quartieri più svantaggiati della capitale del Regno Unito, caratterizzato da molteplici problematiche associate alla privazione economica e sociale. L'amministrazione locale ha associato la necessità di intervenire in modo mirato e preventivo alla crescente digitalizzazione attraverso vari programmi, come il “One View”, una piattaforma di gestione dati che utilizza l'analisi predittiva per fornire interventi mirati a persone vulnerabili (Gentile, 2024; Institute of Global Health Innovation - IGHI, 2021). La piattaforma integra set di dati (precedentemente scollegati), riguardanti i vari servizi, fornendo agli assistenti sociali una visione unificata delle famiglie e degli individui; inoltre, utilizza modelli analitici avanzati per identificare le persone a rischio, segnalando potenziali problematiche con sei-nove mesi di anticipo e permettendo così interventi proattivi (IGHI, 2021). L'integrazione dell'IA in questo contesto non solo ha ottimizzato le risorse disponibili, ma ha anche reso il supporto ai cittadini in difficoltà più proattivo e mirato, riducendo le situazioni emergenziali. A tal fine, il sito web del quartiere¹ offre una panoramica chiara e accessibile dei servizi e delle iniziative promosse dall'amministrazione a beneficio dei residenti. La sezione dedicata al welfare include informazioni sui servizi sociali, l'assistenza abitativa, l'educazione e il sostegno alle famiglie e fornisce supporto a chi cerca lavoro attraverso programmi di formazione, consulenza professionale e servizi per l'impiego. Questa piattaforma è un buon esempio di accessibilità diretta tra i cittadini in situazioni di vulnerabilità e i servizi

¹ Welcome to Barking and Dagenham | London Borough of Barking and Dagenham (lbbd.gov.uk).

necessari perché consente di informarsi ed accedere a ciò di cui si ha bisogno navigando tra le diverse sezioni del sito (Gentile, 2024).

3.2 IA per l'automazione e l'efficienza dei servizi

L'uso dell'IA presso la pubblica amministrazione potrebbe aumentare l'efficienza operativa, accelerando i tempi di risposta e permettendo di destinare le risorse umane ad attività di maggiore valore aggiunto. Secondo il Gruppo di Esperti presso la Commissione Europea, i servizi pubblici dovrebbero investire in interfacce conversazionali in grado di soddisfare le esigenze degli utenti in modo continuativo, offrendo un servizio più agile, accessibile e tempestivo attraverso un unico punto di contatto (AI HLEG, 2019b). Un esempio potrebbe essere l'integrazione delle IA nelle piattaforme che connettono gli utenti agli erogatori di servizi attraverso chatbot interattivi, migliorando la connessione tra i fruitori e gli erogatori dei servizi (Scannavini, 2024). In particolare, i chatbot sono «robot software per la conversazione» (Sartor, 2022) e cioè programmi informatici progettati per simulare una conversazione attraverso messaggi testuali o vocali e per rispondere automaticamente a richieste o domande. Il loro impiego in ambito sociale aumenterebbe la capacità di assistere le persone, indirizzandole verso le informazioni o i servizi desiderati e semplificando la compilazione dei moduli in un contesto conversazionale. In tal senso, la piattaforma del borgo londinese, precedentemente citato, rappresenta un ottimo esempio in cui è stato sviluppato un'interfaccia comunicazionale di questo tipo: l'assistente virtuale risponde a domande frequenti, fornisce informazioni sui servizi e guida gli utenti attraverso le procedure necessarie, migliorando l'accessibilità e riducendo i tempi di attesa (Gentile, 2024). In Italia, un esempio virtuoso in tal senso sono i Chatbot delle aziende addestrati a fornire informazioni sul Welfare

Aziendale e su iniziative e agevolazioni a favore dei dipendenti e delle loro famiglie: l'integrazione dell'IA in questo contesto segna un passo avanti verso un futuro in cui le tecnologie digitali miglioreranno significativamente la qualità della vita lavorativa, supportando la conciliazione delle esigenze vita-lavoro. La digitalizzazione delle misure di sostegno aziendale non solo ottimizza i processi interni, ma risponde anche a una crescente domanda di efficienza e personalizzazione da parte degli utenti (Castells, 2009; Benkler, 2006). In un mondo digitale, i clienti si aspettano che le aziende siano disponibili 24 ore su 24, 7 giorni su 7: i chatbot rappresentano un modo semplice ed economico per farlo perché i visitatori che hanno domande o necessità di supporto tecnico possono ricevere l'assistenza che gli serve, nel momento in cui gli serve. Proprio in riferimento alle interazioni uomo-macchina, gli Esperti di IA hanno elaborato una serie di sfide e preoccupazioni: in particolare ritengono che «Gli esseri umani dovrebbero sempre sapere se stanno interagendo direttamente con un altro essere umano o con una macchina e spetta agli operatori del settore dell'IA la responsabilità di comunicarlo in modo affidabile» (AI HLEG, 2019a). Gli operatori dovrebbero pertanto garantire (ad es. tramite clausole di esclusione della responsabilità chiare e trasparenti) che le persone siano consapevoli del fatto che stanno interagendo con un bot o che possano chiedere informazioni e approvare tale interazione. L'automazione dei processi amministrativi tramite l'intelligenza artificiale rappresenta un'opportunità significativa per migliorare l'efficienza e l'efficacia dei servizi sociali (Gentile, 2024). L'implementazione dell'IA consentirebbe di automatizzare la classificazione delle domande di assistenza, analizzando rapidamente grandi volumi di dati e assegnando le richieste alle categorie appropriate in modo preciso e coerente. In ambito socioassistenziale, le piattaforme online potrebbero facilitare la gestione dei casi, migliorando la

comunicazione e la collaborazione tra le équipes di professionisti e le organizzazioni coinvolte. Ad esempio, sistemi di gestione dei casi basati su cloud consentono agli assistenti sociali di accedere facilmente alle informazioni necessarie, condividere documenti in tempo reale e coordinare interventi in modo più fluido ed efficace. In particolare, molti studi riportano l'uso dell'IA nella pianificazione della distribuzione delle risorse e nella gestione delle richieste di prestazioni, sia in denaro che in natura (Molinuevo, 2020): ciò consente una migliore allocazione delle risorse disponibili, riducendo gli sprechi e assicurando che le necessità vengano soddisfatte in modo tempestivo e preciso. Oltretutto, nell'ambito della pubblica amministrazione e degli enti locali, si potrebbe considerare l'impiego di chatbot generativi per istituire una sorta di vice-segretariato sociale, al quale i destinatari dei servizi potrebbero rivolgersi prima di recarsi di persona a richiedere assistenza. Questo utilizzo consentirebbe di fornire un primo livello di supporto e informazioni, migliorando l'efficienza del servizio e riducendo il carico di lavoro per il personale umano. I chatbot potrebbero aiutare a filtrare le richieste, indirizzando le persone verso le risorse più appropriate e fornendo risposte immediate alle domande frequenti. In tal modo, si ottimizzerebbe l'accesso ai servizi, facilitando un'interazione preliminare che potrebbe risolvere molte esigenze senza la necessità di un incontro faccia a faccia. Buona parte della letteratura sociologica suggerisce che tali strumenti tecnologici possono contribuire a ridurre il digital divide all'interno delle organizzazioni, promuovendo una maggiore inclusione digitale e una distribuzione più equa delle risorse informative (Di Maggio & Hargittai, 2001). L'implementazione di questo tipo di strumenti potrebbe costituire un passo fondamentale verso l'applicazione concreta dei principi di quel Nuovo Welfare – individuato da Ferrera (1996) – che si concentra in

particolare sulle esigenze individuali e su un approccio più personalizzato alla fornitura dei servizi.

3.3 IA per il supporto alla salute e al benessere

Negli ultimi anni, il settore della cura ha dovuto affrontare sfide importanti, come l'aumento della spesa sanitaria e la crescente domanda di assistenza, dovuta all'invecchiamento della popolazione (Istat, 2023). In questo contesto, le tecnologie digitali emergono come una soluzione promettente, in grado di trasformare le modalità di assistenza e migliorare la qualità della vita dei pazienti. Il termine «Care technology» o «Tele Care» è utilizzato proprio per descrivere dispositivi digitali in grado di fornire servizi e assistenza alle persone fragili: tale servizio include sensori personali e ambientali in casa che, grazie all'impiego di algoritmi di IA, monitorano la situazione h24 e che, in caso di imprevisti, trasmettono la comunicazione in tempo reale ai contatti di emergenza (The Voluntary Organisations Disability Group, 2013). Questi strumenti possono significativamente supportare l'indipendenza delle persone non autosufficienti in vari modi: monitorare a distanza le terapie e i parametri vitali dell'assistito, ricordare l'orario delle medicine, segnalare se non sono state assunte, o fornire aiuto nelle faccende domestiche (Scannavini, 2024). Con il termine «Smart home technology» si è soliti riferirsi a case in cui vari componenti sono interconnessi e comunicano tra di loro attraverso una rete (Billings *et al.*, 2013). Grazie a queste tecnologie, le persone non autosufficienti possono vivere in modo più autonomo nelle proprie abitazioni, riducendo la necessità di ricorrere a strutture di assistenza specializzata e i relativi costi. In situazioni di emergenza, i sistemi domestici automatizzati possono contattare immediatamente i caregiver o i familiari, fornendo un ulteriore livello di sicurezza e tranquillità. Questo non solo migliora

la qualità della vita degli assistiti, ma rappresenta anche un risparmio economico significativo per le famiglie e il sistema sanitario. L'utilizzo delle tecnologie informatiche è parte integrante della nostra vita quotidiana e spesso anche gli strumenti che consideriamo semplici gadget domestici trovano applicazione in contesti di cura, dimostrando la loro utilità non solo per l'intrattenimento o la gestione personale, ma anche per la sicurezza e il supporto degli individui più vulnerabili (Ainsworth, 2023). Tra gli oggetti elettronici più diffusi e facili da usare nella vita quotidiana, spiccano gli smartwatch e gli assistenti virtuali dotati di IA che offrono un supporto essenziale, perché permettono di richiedere assistenza in qualsiasi momento e di trasmettere la posizione in tempo reale agli operatori o ai contatti di emergenza.

3.4. IA e inclusione sociale di persone fragili

L'intelligenza artificiale è anche un potente strumento per favorire l'inclusione scolastica e lavorativa delle persone in condizioni di fragilità. In particolare, questi strumenti si rivelano molto utili per le persone anziane perché consentono di mitigare gli effetti dell'isolamento e della solitudine, mantenendo connessioni sociali attive. In tal senso, una ricerca condotta nel 2014, *Cyberseniors and Quality of Life: A Focus on Social Networking*, ha dimostrato come l'utilizzo delle tecnologie di connessione interpersonale migliorino la qualità della vita delle persone anziane, combattendo l'isolamento e migliorandone l'attività cerebrale (Rocha Machado *et al.*, 2014). Le nuove tecnologie e la robotica sono in grado di influenzare profondamente il modo in cui comunichiamo e interagiamo con il mondo: in questo contesto si inserisce la possibilità di impiego di *robot* per fornire compagnia agli anziani e alle persone fragili. Questi robot, dotati di intelligenza artificiale e progettati per interagire con gli esseri

umani, offrono non solo un supporto pratico, ma anche un sostegno emotivo. La loro capacità di adattarsi alle esigenze individuali e di seguire comportamenti sociali specifici consente di creare un ambiente in cui la compagnia è più accessibile, contribuendo a ridurre il senso di isolamento e solitudine. In ambito socioassistenziale, il loro impiego è diversificato: robot che forniscono assistenza logistica, robot sociali o da compagnia, dispositivi con funzioni motivazionali (ricordare la dieta, i medicinali o di fare esercizio fisico) e strumenti per formare e implementare le competenze comunicazionali dei bambini con bisogni speciali (Dahl e Boulos, 2013). Un robot sociale, in particolare, è un robot autonomo o semi-autonomo, dotato di IA e allenato attraverso sequenze algoritmiche, capace di interagire e comunicare con gli esseri umani seguendo comportamenti sociali e regole legate al suo ruolo specifico (Bensi, 2021). La crescente integrazione di questi strumenti nella nostra vita quotidiana sta trasformando profondamente il concetto di compagnia e interazione umana, soprattutto per le persone anziane e quelle sole. Queste nuove modalità di comunicazione e di partecipazione sociale non riguardano solo la tecnologia in sé, ma anche la motivazione, le emozioni e il linguaggio che vengono trasmesse. In questo contesto, ha perfettamente senso affermare che la comprensione della stessa società digitale contemporanea presuppone un interesse profondo verso le nuove forme di interazione e di partecipazione sociale che caratterizzano i nuovi media (D'Ambrosio, 2019). Un altro aspetto interessante è il ruolo della digitalizzazione nel supportare l'inclusione lavorativa delle persone con disabilità o altre fragilità (Rosa, 2023). In particolare, l'erogazione di tecnologie assistive potenziate dall'intelligenza artificiale, può incidere in modo significativo sull'autostima e sull'autonomia individuale (Gentile, 2024). Tali sistemi permettono di superare limitazioni fisiche e psicologiche, consentendo alle

persone di svolgere le proprie mansioni e migliorando la loro competitività nel mercato del lavoro. Diversi progetti europei mirano a migliorare l'accessibilità tecnologica e l'inclusione lavorativa delle persone con disabilità attraverso soluzioni innovative per ridurre le barriere alla loro partecipazione. Il Veritas Project², per esempio, è dedicato allo sviluppo di strumenti finalizzati a promuovere l'autonomia delle persone con disabilità, consentendo loro di gestire le attività quotidiane in modo indipendente, di ampliare le proprie capacità funzionali e di condurre un'esistenza più soddisfacente e attiva (Kuhlmann, 2023). Inoltre, tali tecnologie garantiscono l'accesso all'istruzione e alla formazione anche per coloro che non possono frequentare fisicamente scuole o università attraverso piattaforme di apprendimento online. Le stesse tecnologie offrono, infine, supporto nell'accesso e nello svolgimento delle attività lavorative, rendendo praticabile il telelavoro o l'utilizzo di software specifici che facilitano l'esecuzione delle mansioni professionali (Kuhlmann, 2023). Le opportunità per i fornitori di servizi di tecnologia assistiva di sviluppare soluzioni personalizzate sono ampie, ma ci sono diverse sfide, tra cui il costo elevato di questi strumenti e la scarsa consapevolezza del pubblico. Pertanto, per sfruttare a pieno le opportunità disponibili, è essenziale aumentare gli investimenti, migliorare la pianificazione dei percorsi di cura e promuovere una maggiore consapevolezza e comprensione della tecnologia assistiva tra i cittadini (Voluntary Organisations Disability Group and the National Care Forum, 2013). Dopo aver esaminato i molteplici usi dell'IA nel settore socioassistenziale e averne esplorato i benefici in termini di efficienza e supporto alle pratiche assistenziali, il paragrafo 5 sarà dedicato interamente a un'analisi critica delle sue contraddizioni, evidenziandone i limiti, le implicazioni etiche e i potenziali rischi.

² <https://www.veritas-project.eu/>.

4. Applicazione dell'IA nei servizi sociali in Italia: stato dell'arte e prospettive future

In Italia, l'applicazione dell'IA nei servizi sociali e nel welfare è ancora limitata, anche se esistono alcune recenti sperimentazioni sia per quanto riguarda i processi amministrativi, sia in ambito sociosanitario.

In particolare, la digitalizzazione sta trasformando profondamente la pubblica amministrazione, migliorando l'accesso ai servizi sociali, l'efficienza dei processi e la trasparenza. Tecnologie come le piattaforme online e le applicazioni mobili permettono ai cittadini italiani di utilizzare strumenti come la CielID (Carta d'Identità Elettronica) e lo SPID (Sistema Pubblico di Identità Digitale) per accedere ai servizi offerti da enti pubblici come i Comuni e l'INPS. Questi strumenti digitali semplificano la fruizione dei servizi, riducono i tempi di attesa e offrono una maggiore sicurezza nell'identificazione e nell'autenticazione degli utenti (AGID, 2021). Un altro esempio è l'utilizzo di GePI, un'applicazione progettata e sviluppata per semplificare il lavoro degli assistenti sociali nell'accompagnare i beneficiari dell'Assegno di Inclusione (ADI) convocati dai servizi sociali dei Comuni. L'applicazione aiuta gli operatori a gestire i beneficiari dell'ADI, facilitando i Patti per l'inclusione sociale e lo scambio di dati tra INPS e Comuni, migliorando così l'efficacia dei programmi di inclusione. Tale piattaforma, pur avendo introdotto significativi miglioramenti nella gestione dei servizi sociali, presenta diverse criticità. Gli utenti hanno segnalato malfunzionamenti tecnici che rallentano le operazioni quotidiane e ostacolano la gestione tempestiva dei casi. Ad esempio, sono numerosi i cittadini bisognosi che, nella transizione dalla tutela del Reddito di Cittadinanza al Reddito di Inclusione, hanno visto sospendere l'erogazione del sostegno economico a causa di problemi nel

trasferimento dei dati tra enti. Questa problematica evidenzia come le inefficienze nella gestione delle informazioni possano avere gravi ripercussioni sui beneficiari dei servizi sociali. L'adozione di questo strumento ha inoltre richiesto un considerevole sforzo formativo per il personale, che ha dovuto familiarizzare con nuove procedure e strumenti. Il Ministero del Lavoro e delle Politiche Sociali ha promosso interventi formativi gratuiti per uniformare e migliorare le metodologie di intervento a livello nazionale. Questo periodo di adattamento è cruciale per garantire l'uso efficace della piattaforma, però rappresenta anche un onere aggiuntivo e gravoso per le amministrazioni locali. All'interno del nostro Paese, l'impiego dell'intelligenza artificiale nel settore del welfare è frequentemente correlato ai cosiddetti bisogni complessi, ovvero a quelle esigenze che richiedono un'integrazione elevata tra assistenza sociale e sanitaria e necessitano di un approccio multidisciplinare che unisce le competenze di vari professionisti del settore. In Italia, l'invecchiamento della popolazione, l'aumento delle malattie croniche e risorse economiche e umane limitate hanno messo pressione sul sistema sociosanitario, rendendo necessarie riforme per migliorare i servizi di salute, ottimizzare l'assistenza fornita dai professionisti e razionalizzare le risorse economiche (Cingolani *et al.*, 2023). L'IA offre potenzialità significative in questo contesto, consentendo una gestione più efficace delle informazioni, il monitoraggio delle condizioni di salute e la personalizzazione dei trattamenti. Nell'Agenda Digitale del Veneto 2020, i servizi sociali sono stati identificati come uno dei dieci settori tematici prioritari, evidenziando così un impegno concreto nel migliorare l'efficienza e l'accessibilità dei servizi alla popolazione attraverso l'adozione di soluzioni tecnologiche avanzate. Per molti, il Veneto rappresenta un esempio virtuoso di sviluppo digitale ed è in grado di offrire un modello replicabile per altre regioni italiane: questa strategia è il risultato di un approccio "bottom-up" attraverso l'ascolto del

territorio tramite incontri tematici (Potti, 2017). Si prevedeva, inoltre, la promozione di dispositivi indossabili, sensori domestici e teleassistenza per sostenere l'invecchiamento attivo con l'obiettivo di creare una rete coordinata di progetti in cui la componente digitale migliori l'autonomia e la qualità della vita delle persone anziane (Molinuevo, 2020). Passando all'Agenda Digitale del Veneto 2025, si nota un cambiamento significativo: il concetto di "sanità sociale" sostituisce quello di "servizi sociali". L'adozione del termine "sanità sociale" evidenzia un impegno crescente nella digitalizzazione e nell'integrazione dei servizi sociosanitari e suggerisce un approccio più olistico e integrato, dove i confini tra salute e servizi sociali diventano più fluidi, con l'obiettivo di fornire un'assistenza più completa e coordinata ai cittadini. L'integrazione sociosanitaria può portare a una migliore gestione delle risorse, un miglioramento nella continuità delle cure e una risposta più tempestiva e personalizzata alle necessità dei cittadini. Recentemente, la regione Veneto è stata presa come esempio di successo anche nel campo degli Open Data, grazie a una visione particolarmente lungimirante: la regione ha avviato diverse iniziative supportate da finanziamenti europei, affrontando principalmente gli aspetti tecnici e le sfide legate agli Open Data, e ha poi proseguito rafforzando il loro utilizzo per supportare la crescita locale e la trasformazione urbana (Interreg Europe, 2024). Il rapporto sulla digitalizzazione realizzato dalla Commissione Europea nei vari paesi dell'Unione Europea ha evidenziato che l'Italia dedica il 25,6% del suo Piano Nazionale di Ripresa e Resilienza alla digitalizzazione, con un totale di 47 miliardi di euro: questa cifra rappresenta un'importante opportunità, ma non è sufficiente per raggiungere completamente gli obiettivi del Decennio Digitale (European Commission, 2024). Il Rapporto dell'Osservatorio per l'Intelligenza Artificiale (2024) indica che, nel periodo 2022-2023, indica che l'Italia ha fatto progressi nella formazione di ricercatori

in questo settore, grazie all'istituzione di nuove borse di dottorato e all'aumento dei finanziamenti per le posizioni di ricerca; tuttavia, continua a faticare nell'attrarre e trattenere talenti, con il 69% delle principali organizzazioni italiane che considera le aziende estere più attraenti. Inoltre, grazie al sostegno del Recovery and Resilience Facility, l'Italia ha intrapreso numerose iniziative per aumentare il numero di servizi integrati con SPID e CIE e incentivare il loro utilizzo. Nel 2023, gli SPID attivi in Italia hanno superato i 37 milioni e le CIE i 39 milioni, con circa il 40% della popolazione che ha utilizzato questi strumenti per accedere ai servizi pubblici online, superando la media UE del 36% (European Commission, 2024). Nonostante il basso livello di competenze digitali nel paese, questi dati rappresentano un significativo progresso. La Commissione Europea ha riconosciuto l'impegno dell'Italia nella digitalizzazione dei servizi pubblici, anche grazie agli ingenti investimenti pervenuti in questo settore: in particolare, il nuovo Piano Digitale per la pubblica amministrazione, stabilito nel 2024, riafferma l'obiettivo di costruire piattaforme più efficienti ed accessibili per i cittadini.

5. Rischi e difficoltà legati all'impiego dell'IA nel settore del welfare

Coerentemente con i piani strategici europei, la “*Strategia Italiana per l'Intelligenza Artificiale 2024-2026*”, promossa dall'Agenzia per l'Italia Digitale (AGID), si propone di orientare lo sviluppo dell'IA non solo verso l'innovazione tecnologica, ma anche come strumento per il benessere sociale degli italiani al fine di apportare importanti cambiamenti nel sistema del welfare. Tuttavia, l'avanzamento dell'intelligenza artificiale comporta una serie di questioni e problematiche complesse che necessitano di

una riflessione approfondita e di un'analisi accurata. In primo luogo, all'interno del tema delle responsabilità per le decisioni prese da sistemi autonomi rientra il problema significativo e complesso dei *bias algoritmici* nel training delle IA: questo fenomeno si verifica quando i modelli riflettono o amplificano i pregiudizi esistenti nei dati utilizzati per il loro addestramento. È evidente che, se i dati di addestramento non rappresentano adeguatamente la popolazione generale, il modello può generare pregiudizi (O'Neil, 2017). Se i dataset utilizzati per addestrare i modelli decisionali non sono rappresentativi o sono contaminati da stereotipi, l'IA potrebbe perpetuare ed enfatizzare ingiustizie, penalizzando gruppi vulnerabili come minoranze etniche, persone con disabilità o cittadini a basso reddito (Lazzini, 2023). Sono diversi i casi conclamati in cui illustri studiosi hanno potuto verificare il perpetuarsi di decisioni viziate da bias discriminatori interiorizzati dagli algoritmi di IA in diversi settori. Ad esempio, gli algoritmi di valutazione del credito possono discriminare contro minoranze etniche o gruppi socio-economicamente svantaggiati. I modelli di IA utilizzati nelle decisioni di concessione di prestiti possono avere un bias sistematico contro i richiedenti di colore o contro coloro provengono da determinate aree geografiche, portando a tassi di approvazione più bassi e a tassi di interesse più alti per questi individui (Bagni, 2021). Anche gli algoritmi di rischio usati nel sistema di giustizia penale possono avere pregiudizi razziali, sviluppati sulla base delle precedenti sentenze raccolte negli archivi dei tribunali. In riferimento a tale aspetto, uno studio di ProPublica (2016) ha mostrato che il sistema COMPAS, utilizzato per valutare il rischio di recidiva negli USA, tendeva a classificare – erroneamente – le persone di origine afroamericana come ad alto rischio più frequentemente rispetto a quelle di origine caucasica (Anwing *et al.*, 2016). Questo tipo di bias può avere gravi conseguenze, influenzando le decisioni dei giudici relativamente a

condanne, cauzioni, libertà condizionali e altre misure. In ambito sociosanitario, lo studio di Obermeyer *et al.* (2019) ha messo in luce il bias razziale insito in un algoritmo utilizzato per la gestione della salute negli USA. Il sistema, apparentemente neutrale, assegna le risorse sanitarie basandosi sulla spesa sanitaria passata, presupponendo che chi ha speso di più in cure abbia maggior bisogno di assistenza futura. Tuttavia, poiché i pazienti afroamericani, a parità di condizioni mediche, tendono a spendere meno rispetto ai bianchi, spesso a causa di ostacoli economici, l'algoritmo sottovaluta sistematicamente i loro bisogni, riducendo del 46% il numero di pazienti che ricevono accesso a cure adeguate (Obermeyer *et al.*, 2019). Questi casi dimostrano come gli algoritmi non siano neutrali, ma possano riflettere e amplificare le disuguaglianze preesistenti, evidenziando la necessità di una progettazione etica e consapevole dell'intelligenza artificiale e delle variabili di input, soprattutto nei settori in cui le decisioni algoritmiche incidono direttamente sui diritti e sul benessere delle persone. Approfondendo il tema delle decisioni prese dai sistemi autonomi, alcuni studiosi, come Lazzini (2023), sottolineano come il *decisionismo algoritmico* possa condurre a esiti distorti, ad esempio nella valutazione della necessità di assistenza sociale o nell'assegnazione di sussidi. I sistemi automatizzati potrebbero escludere richiedenti che in realtà avrebbero diritto a determinati benefici, solo perché il loro profilo non rientra nelle categorie previste dagli algoritmi. In questo modo, l'IA potrebbe rafforzare logiche di esclusione piuttosto che promuovere un accesso equo ai servizi. Le decisioni automatizzate, utilizzate per risparmiare e semplificare i processi, spesso risultano discriminatorie, negando assistenza a chi ne ha bisogno: la Eubanks (2018) descrive questo fenomeno come un "*digital poorhouse*", un sistema in cui i più vulnerabili sono ulteriormente penalizzati per l'accesso ai servizi essenziali, perpetuando cicli di povertà e marginalizzazione. Un

ulteriore problema riguarda la *mancanza di trasparenza* nei processi decisionali basati sull'IA. Molti algoritmi utilizzati nel welfare funzionano come “scatole nere”, ossia senza che i cittadini e gli operatori sociali possano comprendere pienamente i criteri utilizzati per prendere decisioni. Questo crea un *deficit di responsabilità*, poiché diventa difficile contestare decisioni ingiuste o errate. Lazzini (2023) sottolinea che, in un sistema democratico, è fondamentale garantire *meccanismi di controllo e revisione umana* per evitare che l'IA prenda decisioni arbitrarie o errate senza possibilità di appello. Senza adeguate regolamentazioni e principi etici, il rischio è che il welfare diventi sempre più disumanizzato, trasformando i cittadini in semplici numeri all'interno di un sistema automatizzato. Dal punto di vista sociologico, gli strumenti di IA hanno il potenziale di modificare le strutture di potere (che ad oggi riguardano prioritariamente la distribuzione delle risorse economiche) e inasprire le disuguaglianze sociali. L'adozione dell'IA in vari settori può portare a una redistribuzione del potere, spesso favorendo coloro che già possiedono le risorse e le competenze necessarie per sfruttare i dispositivi informatici. Le tecnologie avanzate richiedono infrastrutture digitali e competenze specifiche, ma l'accesso a queste risorse non è uniforme tra regioni e comunità. Le aree con limitata connessione internet o poche risorse tecnologiche rischiano di rimanere indietro, aumentando il divario tra chi può sfruttare appieno le potenzialità dell'IA e chi no. In una “società interconnessa” l'esclusione digitale è una delle principali sfide nell'implementazione dell'intelligenza artificiale nel settore socioassistenziale perché rischia di isolare chi non possiede dispositivi elettronici e di contribuire all'aumento delle disuguaglianze (Dominici, 2018; Parra-Saiani, 2020). Ciò potrebbe configurare anche il rischio di esclusione sociale alla cura, limitando l'accesso a servizi essenziali basati su tecnologie intelligenti, come la prenotazione di visite, l'assistenza domiciliare

digitale e i sistemi di supporto personalizzati. Senza adeguate strategie di inclusione, gli anziani e le fasce più vulnerabili della popolazione potrebbero trovarsi ulteriormente emarginati, privati delle opportunità offerte dall'IA per migliorare la qualità della vita e il benessere sociale. Un altro aspetto da tenere in considerazione in ambito socioassistenziale è il problema della tutela della *privacy* e della sicurezza dei dati personali delle persone che si rivolgono ai servizi. L'utilizzo dell'IA nell'ambito di un settore complesso come quello socioassistenziale implica spesso la necessità di analizzare grandi quantità di informazioni sensibili, tra cui dati finanziari, sanitari e sociali. Lazzini (2023) evidenzia il rischio di *sorveglianza digitale*, in cui gli individui vengono monitorati e classificati sulla base di criteri automatizzati, con un impatto significativo sulla loro libertà e dignità. In alcuni casi, governi e istituzioni potrebbero utilizzare l'IA per controllare e limitare l'accesso ai servizi assistenziali, introducendo meccanismi di valutazione predittiva che anticipano il comportamento futuro degli utenti. Ad esempio, sistemi predittivi potrebbero penalizzare persone ritenute "a rischio" di frode o di comportamenti non conformi, anche senza prove concrete, creando una forma di *discriminazione preventiva* (Lazzini, 2023). In tal senso, gli assistenti sociali e, in generale, gli operatori del sociale sono chiamati ad affrontare la delicata questione di raccogliere e gestire le informazioni sensibili in modo tale che i diritti degli individui siano sempre tutelati. In questo discorso si inserisce anche il tema della *vulnerabilità dei dispositivi di raccolta dati e della cybersecurity*, soprattutto in relazione al rischio di attacchi informatici. La concentrazione di enormi quantità di dati personali nelle mani di enti pubblici e del privato sociale solleva interrogativi sulla sicurezza informatica e sull'eventuale utilizzo improprio delle informazioni raccolte. Il pericolo di fughe di dati, accessi non autorizzati o manipolazioni delle informazioni potrebbe compromettere gravemente la

riservatezza dei cittadini più vulnerabili, esponendoli a ulteriori forme di sfruttamento e marginalizzazione. Per affrontare questa problematica, è fondamentale che gli operatori del settore socioassistenziale non solo acquisiscano competenze nell'uso delle tecnologie digitali, ma siano anche adeguatamente formati per garantire la protezione e la sicurezza dei dati delle persone assistite. Inoltre, diventa indispensabile adottare soluzioni di sicurezza avanzate, capaci di prevenire minacce informatiche, accessi non autorizzati e manipolazioni dei dati. Parallelamente, i servizi pubblici che trattano dati sensibili, in qualsiasi settore del welfare, dovrebbero essere supportati da enti specializzati in sicurezza informatica in grado di garantire una protezione adeguata e un costante monitoraggio delle infrastrutture digitali. Solo un approccio integrato tra tecnologia, formazione e sicurezza può assicurare che l'intelligenza artificiale venga impiegata in modo realmente sicuro e rispettoso dei diritti delle persone.

Conclusioni

Sin dai tempi della Rivoluzione Industriale, gli sviluppi significativi nell'innovazione tecnologica hanno avuto un impatto profondo nel trasformare numerosi compiti manuali e processi che erano stati praticati per decenni, spesso quando gli esseri umani avevano raggiunto i limiti della propria capacità fisica; l'IA possiede lo stesso potenziale trasformativo, sia per l'ottimizzazione che per la potenziale sostituzione delle attività e dei compiti umani in una vasta gamma di applicazioni industriali, intellettuali e sociali (AA.VV., 2021). Dal punto di vista sociologico, il vasto e pervasivo impiego dell'IA può essere interpretato attraverso le lenti del concetto di "paradigma" elaborato da T. Kuhn. Secondo Kuhn (1962), il progresso scientifico non è lineare, ma avviene attraverso

una serie di rivoluzioni che portano a cambiamenti fondamentali nei modi in cui comprendiamo e interagiamo con il mondo. La transizione verso l'IA rappresenta un nuovo paradigma, in cui la tecnologia non solo aumenta le capacità umane, ma ridefinisce anche il concetto stesso di lavoro e produttività, sollevando questioni inedite. Così, l'integrazione dell'IA nelle nostre vite quotidiane non è solo un'innovazione tecnologica, ma un fenomeno complesso e rivoluzionario che richiede un'analisi critica e una riflessione sulle sue implicazioni per la società nel suo complesso. In questo nuovo contesto, diventa necessario un intervento più incisivo della Sociologia, per dare il proprio contributo unico e indispensabile all'interpretazione e alla formazione di competenze specifiche necessarie per leggere, comprendere, creare e comunicare dati per orientare il pensiero verso la priorità della comprensione dei fenomeni sociali (Capogna, 2022). L'intelligenza artificiale potrebbe rappresentare una risorsa straordinaria per innovare e migliorare i servizi socioassistenziali in Italia, offrendo strumenti più efficaci per rispondere alle esigenze dei cittadini. Tuttavia, per massimizzare questi vantaggi, è cruciale affrontare con rigore le sfide inerenti alla protezione della privacy, ai bias algoritmici e all'accesso equo all'informatizzazione. Emerge chiaramente la necessità di identificare soluzioni e strategie per massimizzare i benefici e minimizzare l'impatto sociale negativo delle nuove tecnologie nel welfare. In primo luogo, è fondamentale sviluppare una profonda consapevolezza dei limiti e dei rischi associati alla loro implementazione. Questo richiede un'analisi approfondita e continua delle implicazioni etiche, sociali ed economiche delle tecnologie digitali applicate ai servizi di welfare, anche attraverso la conduzione di studi di caso nei settori specifici. Un'altra possibile soluzione risiede nella definizione di *linee guida chiare ed efficaci*, che includano le *Best Practices*. Tali linee guida devono essere concepite per governare un sistema complesso e

spesso transnazionale, capace di adattarsi alle diverse realtà locali pur mantenendo standard elevati di efficienza e equità. Dovrebbero fornire orientamenti dettagliati su come utilizzare le infrastrutture computazionali per migliorare l'accesso ai servizi, proteggere i dati personali e garantire che nessun gruppo sociale venga escluso. Inoltre, è cruciale *sensibilizzare la cittadinanza* sull'importanza e l'utilizzo delle nuove tecnologie nel welfare. Questo può essere realizzato attraverso campagne di informazione e formazione che spieghino i benefici e i rischi associati all'uso di piattaforme digitali e risorse informatiche nei servizi sociali. La partecipazione attiva dei cittadini nel dibattito su queste tematiche può contribuire a creare un clima di fiducia e collaborazione tra la popolazione e le istituzioni. Altro aspetto essenziale è l'aggiornamento continuo degli operatori sociali, affinché siano sempre informati sulle nuove tecnologie e le migliori pratiche per integrarle nel loro lavoro quotidiano. La formazione dovrebbe includere non solo competenze tecniche, ma anche aspetti etici e relazionali, per garantire che l'uso di tali strumenti non comprometta la qualità del servizio e il rapporto umano con le persone. In questo modo, gli operatori saranno in grado di utilizzare efficacemente le nuove tecnologie, migliorando l'efficienza e la qualità dei servizi offerti.

In conclusione, per affrontare efficacemente le sfide poste dall'integrazione delle nuove infrastrutture digitali nel welfare, è necessario adottare un approccio multidimensionale che includa queste proposte in modo sinergico e ragionato. Ciò è possibile solo attraverso la collaborazione tra gli attori istituzionali, gli enti del terzo settore, la comunità accademica e la società civile. Al tempo stesso, come sottolinea Dominici (2018), è importante che tale innovazione sia supportata da una cultura della comunicazione e da una profonda comprensione della complessità. Sarebbe opportuno promuovere politiche sociali illuminate e a lungo termine, focalizzandosi in particolare su scuole e università, per

favorire e mantenere un cambiamento culturale significativo. L'inclusione deve essere una priorità, affinché nessuno venga lasciato indietro, e ciò implica investimenti in formazione e alfabetizzazione digitale, accanto a politiche che vadano ad incentivare l'uso delle tecnologie in modo responsabile e consapevole. In questo contesto, un sistema di welfare efficace è quello che non solo si adatta ai cambiamenti tecnologici, ma li integra in modo strategico per migliorare l'accesso ai servizi sociali, rendere più efficiente la distribuzione delle risorse e facilitare l'interazione tra cittadini e istituzioni.

Bibliografia

AA.VV. (2021, aprile). *Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities and agenda for research, practice and policy*, International Journal of Information Management, Vol. 57. <https://doi.org/10.1016/j.ijinfomgt.2019.08.002>.

Agenzia per l'Italia Digitale (2024; 2021). *Strategia Italiana per l'Intelligenza Artificiale 2024-2026 e Servizi pubblici online: si accede con SPID e CIE*.

Ainsworth, A. (2023, 25 maggio). *How can digital technologies support care?*, UK. <https://socialcare.blog.gov.uk/2023/05/25/how-can-digital-technologies-support-care/>.

Alston, P. (2019, 11 ottobre). *Report of the Special Rapporteur on Extreme Poverty and Human Rights*, United Nations General Assembly A/74/493.

Angwin J., Larson J., Mattu S. e Kirchner L. (2016). *Pregiudizio sulle norme*, ProPublica. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

Ardigò, A. (1988). *Per una sociologia oltre il post-moderno*. Roma: Laterza.

Bagni, F. (2021). Uso degli algoritmi nel mercato del credito: dimensione nazionale ed europea. *Osservatorio sulle fonti*, 2. <http://www.osservatoriosullefonti.it>.

Benanti, P. (2018). *Homo Faber. The Techno-Human condition*, EDB - Edizioni Dehoniane Bologna.

Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven: Yale University Press.

Bensi, T. (2021). I Robot Sociali e Assistenziali potrebbero aiutare anche te, *Software Uno*. 15 marzo.

Billings, J. et al. (2013). *The role of information technology in long-term care for older people*. In *Long-Term Care in Europe: Improving Policy and Practice*, London: Palgrave Macmillan UK. 252-277.

Capogna, S. (2023). Sociology between big data and research frontiers: A challenge for educational policies and skills. *Quality & Quantity*, 57(1). 193–212. <https://doi.org/10.1007/s11135-022-01351-7>.

Castells, M. (2009). *The rise of the network society: The information age: Economy, society, and culture*. Wiley-Blackwell.

Cingolani, M. *et al.* (2023, 5 gennaio). Artificial intelligence and digital medicine for integrated home care services in Italy: Opportunities and limits. *Frontiers in Public Health*, 10 - 2022. <https://doi.org/10.3389/fpubh.2022.109500>.

Cipriani, R., Pace, E., & Losito, G. (2020). *L'incerta fede: Un'indagine quanti-qualitativa in Italia*. Milano: Franco Angeli.

Commissione Europea, (2018, 25 aprile). *Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato Economico e Sociale Europeo e al Comitato delle Regioni. L'intelligenza artificiale per l'Europa*, Bruxelles.

Commissione europea (2021, 21 aprile), *Proposta di regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza Artificiale (Legge sull'Intelligenza Artificiale) e modifica alcuni atti legislativi dell'Unione*, Bruxelles.

Commissione Europea (2024, 2 luglio). *Digital Decade 2024: Country reports: Italy*, Policy and legislation, Bruxelles.

Consiglio UE (2024). *Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union Legislative acts* – Analysis of the final compromise text with a view to agreement, Bruxelles.

Corradini, F., Fedeli, A., Fornari, F., Polini, A., & Re, B. (2021). *FloWare: an approach for IoT support and application development. In International Conference on Business Process Modeling, Development and Support*. Cham: Springer International Publishing.

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.

Dahl, T. S. and Boulos, M. N. K. (2013). Robots in health and social care: A complementary technology to home care and tele healthcare? *Robotics*, 3(1), 1-21.

Derbring S, Barbos Nordström M, Svenningsson J-A, *et al.* (2023). Effects of a digital reminiscing intervention on people with dementia and their care-givers and relatives. *Ageing and Society*, 43(9):1983-2000. doi:10.1017/S0144686X21001446.

Di Maggio, P., & Hargittai, E. (2001). *From the ‘Digital Divide’ to ‘Digital Inequality’: Studying Internet Use as Penetration Increases*, Princeton University, Woodrow Wilson School of Public and International Affairs, Center for Arts and Cultural Policy Studies, Working Papers.

Dipartimento per la trasformazione digitale (2022, 5 settembre). *Identità digitale: L’accesso semplice e veloce ai servizi digitali*. <https://innovazione.gov.it/progetti/identita-digitale-spide-cie/>.

Dominici, P. (2018). For an inclusive innovation. Healing the fracture between the human and the technological in the hypercomplex society. *European Journal of Futures Research*, 6(1), 1-10.

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

Ferrera, M. (1996). The “Southern Model” of Welfare in Social Europe. *Journal of European Social Policy*, 6(1), 17-37. <https://doi.org/10.1177/095892879600600102>.

Gentile, M. (2024, 24 settembre). *Tecnologie di IA nei servizi sociali e sanitari. Vantaggi e Considerazioni*, Welforum. <https://www.welforum.it/tecnologie-di-ia-nei-servizi-sociali-e-sanitari/>.

Gruppo indipendente di esperti ad alto livello sull'intelligenza artificiale (High-Level Expert Group on AI - AI HLEG) (2019a; 2019b). *Ethics guidelines for trustworthy AI e Policy and investment recommendations for trustworthy Artificial Intelligence*. <https://digital-strategy.ec.europa.eu/en/library>.

Hatzius, J. (2023). *The Potentially Large Effects of Artificial Intelligence on Economic Growth* (Briggs/Kodnani), Goldman Sachs.

Institute of Global Health Innovation (2021). *One View: London Borough of Barking and Dagenham*. Imperial College London, London, UK.

Interreg Europe, (2024, 14 maggio). Open Data: Spotlight on Veneto Region. <https://www.interregeurope.eu/find-policy-solutions/stories/open-data-spotlight-on-veneto-region>.

Istat (2023). *Rapporto Annuale 2023. La situazione del Paese*, Roma: Marchesi Grafiche Editoriali.

Kissinger, H. A. et al. (2023). *L'erA dell'Intelligenza artificiale. Il futuro dell'identità umana*. Milano: Mondadori.

Kuhlmann, R. (2023). *Assistive technologies for people with disabilities*, <https://www.veritas-project.eu/assistive-technologien-fuer-menschen-mit-behinderungen/>.

Kuhn, T. S. (2009). *La struttura delle rivoluzioni scientifiche* (3^a ed.). Milano: Einaudi (Original work published 1962).

Lazzini, F. (2022). *Etica digitale e Intelligenza Artificiale. I rischi per la protezione dei dati*. Torino: Giappichelli.

McFarland A. (2024, 1 ottobre). *I 9 migliori strumenti di intelligenza artificiale per analisti di dati*. Unite AI.

Ministero del Lavoro e delle Politiche Sociali, (2021, 24 settembre). *Al via la quarta edizione del corso per Case Manager Rdc*, <https://www.lavoro.gov.it/redditodicittadinanza/Formazione/Pagine/default>.

Molinuevo, D. (2020). *Public services Impact of digitalisation on social service*, European Foundation for the Improvement of Living and Working Conditions. <http://eurofound.link/ef19043>.

Moon, S. & Park, K. (2020). The effect of digital reminiscence therapy on people with dementia: a pilot randomized controlled trial, *BMC Geriatrics*, 20, 166. <https://doi.org/10.1186/s12877-020-01563-2>.

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>.

OECD (2019), *Artificial Intelligence in Society*. Paris: OECD Publishing. <https://doi.org/10.1787/eedfee77-en>.

O'Neil, C. (2017). *Armi di distruzione matematica: Come i Big Data aumentano la disuguaglianza e minacciano la democrazia*. Firenze/Milano: Giunti/Bompiani.

Osservatorio per l'Intelligenza Artificiale, (2024. 31 gennaio). *Contributo dell'Osservatorio Artificial Intelligence all'Indagine Conoscitiva sull'intelligenza artificiale: opportunità e rischi per il sistema produttivo italiano*, Camera dei Deputati.

Panesar, A. (2019). *Machine Learning and AI for Healthcare: Big Data for Improved Health Outcomes*. Berkeley, CA: Apress.

Parra-Saiani. P. (2020). Verso un welfare state digitale? L'intelligenza artificiale tra politiche sociali e apparati di controllo. *La Rivista delle Politiche Sociali/ Italian Journal of Social Policy*, 4, 237-254. <https://www.futura-editrice.it/wp-content/uploads/2021/03/RPS-4-2020-13-Parra-Saiani.pdf>.

Potti, G. (2017). *Perché l'Agenda Digitale del Veneto può essere modello per altre Regioni*, ADV2020. <https://www.agenda digitale.eu/infrastrutture/agenda-digitale-veneto-lapproccio-dal-basso-che-la-rende-un-modello/>.

Regolamento Generale sulla protezione dei dati (2016, 27 aprile) Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio.

Rocha Machado L., Jantsch A., Valdeni De Lima J. e Behar P. A. (2014). *Cyberseniors and quality of life: A focus on social networking*, Proceedings of the European Conference in e-Learning, ECEL.

Rosa, D. (2023). *Intelligenza artificiale e robotica al servizio delle persone con disabilità*. Affari Italiani.

Russel, S. & Norvig, P. (2005). *Intelligenza artificiale. Un approccio moderno*, Vol. 1, 2^a Edizione. Milano: Pearson Education Italia.

Sartor, G. (2022). *Intelligenza artificiale e diritto*. Torino: Giappichelli.

Scannavini, M. (2024). *AI e welfare in Italia: cosa cambierà?*, Percorsi di Secondo Welfare <https://www.secondowelfare.it/primo-welfare/ai-e-welfare-in-italia-cosa-cambiera/>.

Unesco (2023). *Global education monitoring report, 2023: technology in education: a tool on whose terms?* <https://doi.org/10.54676/UZQV8501>.

Vetrano, D. L. & Vaccaro, K. (2017). *La babele dell'assistenza domiciliare in Italia: Chi la fa, come si fa*, Roma: Italia Longeva.

Voluntary Organisations Disability Group and the National Care Forum (2013). *Using assistive technology to support personalisation in social care*, London. www.housinglin.org.uk.

World Economic Forum-WEF (2023). *The Future of Jobs Report 2023*. <https://www.weforum.org/publications/the-future-of-jobs-report-2023/>.

Zahidi, S. (2023, 30 aprile). *Prefazione in The Future of Jobs Report 2023*, World Economic Forum.

5. A MODEL FOR RESPONSIBLE GOVERNANCE OF HUMAN-CENTRIC AI IN THE PUBLIC SECTOR

by Francesco Niglia*

Abstract: La governance responsabile dell'Intelligenza Artificiale nel settore pubblico non è più un'opzione rimandabile a causa dei numerosi problemi etici emersi negli ultimi anni caratterizzati da una massiccia adozione di servizi basati sull'IA. Date le numerose sfide poste, diviene obbligatorio includere prospettive sociali incentrate sull'uomo. Questo studio discute un modello di framework per definire i ruoli, le responsabilità e le competenze di tutti gli stakeholder coinvolti nei processi di sviluppo, distribuzione e valutazione dell'IA nel settore pubblico.

Parole chiave: governance, responsabilità, affidabilità, modello, pubblica amministrazione, antropocentrismo.

Abstract: Responsible AI Governance in the Public Sector is no longer an option due to the numerous ethical issues that have emerged in recent years with the adoption of AI-based services in the Public Sector. Given the numerous challenges AI poses, it is essential to incorporate human-centric and social perspectives. This study discusses a framework model for defining the roles, responsibilities, and skills of all the stakeholders involved in the processes of AI development, deployment, and assessment.

* Università degli Studi "Link", <https://orcid.org/0000-0001-6452-5189>, f.niglia@unilink.it.

Accettato Dicembre 2024 - Pubblicato Aprile 2025.

Keywords: governance, responsibility, trustworthiness, framework, public sector, human-centric.

Introduction

This essay builds upon theoretical research to develop a framework for implementing responsible governance of Artificial Intelligence in the public sector and identifying the minimum skills necessary to achieve this goal. The core research question lies in providing evidence about the operational feasibility of EU trustworthy AI guidelines by deeply analysing ethical and responsibility requirements and connecting the dots between the theoretical approaches and early practice of AI governance examples. The study draws on official European Commission reports on the ethical deployment and use of AI and several existing frameworks proposed by renowned research teams and international innovation agencies that aim to implement social and ethical AI. The essay's strategy examines how responsibilities among stakeholders are being reconfigured to achieve new equilibria necessitated by introducing AI-based solutions in the public sector and consequently outlines a possible framework of roles and competencies.

1. What is Responsible Human-Centric AI Governance, and why is it relevant

Artificial Intelligence Governance in the Public Sector is not more of an option since AI has become a «key strategic asset of public service in the last few years» (Tangi L. *et al.*, 2022), and the whole Public Sector witnessed it through its rapid engagement in

the government's AI appropriation procedures (European Commission (2024a)). The capacity of AI technologies and systems to operate in ways that were not previously possible may challenge our existing legal, moral, and social conceptions of responsibility, particularly given their inner properties. The deployment of AI systems has given many examples of how it could impact or infringe on human rights and human-centred values, decreasing or excluding human needs, values, and capabilities. Such characteristics leverage the capacity to generate 'hidden' insight from merging data sets, the ability to accurately imitate human traits, greater software complexity, inscrutability and opacity, and the risk of generating collective action problems. AI ethics is, therefore, required as an integral component due to its potential to touch many aspects of society. Policymakers and all actors in AI should implement safeguards to address significant risks arising from purposeless, intentional, or unintentional misuses. Although AI development has quickly outpaced the speed of regulation, policymakers and regulators have started fining companies for either developing or simply using biased AI algorithms (Hak, A., 2022). As a further confirmation that Ethics and Responsibility in AI development are a global issue, we can list the growing number of international standards highlighting a broad and shared consensus for ensuring the adoption of responsible AI. Starting from the EU AI Act in 2019, examples are the OECD AI Principles and the UNESCO Recommendations on the Ethics of AI in 2021, the G7 Toolkit for AI in the Public Sector, and the UN Resolution in 2024 (UN, 2024). Responsible Human-centric AI Governance (selected acronym RHAIG for this study's purposes) constitutes a common requirement in this international scenario; however, despite being a simple concept, it reveals a pretty complex structure behind it because it harmonises Responsibility, Trustworthiness, and Human-Centric approaches. The first step in

untangling the “conceptual knots” is to clarify the differences among these terms. Responsibility is the process of defining policies and establishing accountability to guide the creation and deployment of AI systems in an organisation. Its scope and criteria have been determined by the High-Level Expert Group on Artificial Intelligence (AI-HLEG, 2019). It leverages four ethical principles and seven requirements (details in Table 1): ethical principles are imperatives rooted in fundamental rights, which must be respected to ensure that AI systems are developed, deployed and used in a trustworthy and human-centric manner; Key Requirements can translate the principles into concrete actions to achieve Trustworthy AI; they include systemic, individual and societal aspects and apply to different stakeholders in AI systems’ life cycle: developers, deployers, end-users, and the broader society. Trustworthiness in AI ensures fairness, inclusion, and benefit for all members of society. It is central to the future acceptance and adoption of AI technologies in all public areas: the realisation of trusted technologies requires cross-cutting collaboration in governance (law, regulation, ethics), technology (explainability, uncertainty, fairness/bias), and end-user application domains (health, business, entertainment, etc.). Human-centred AI prioritises human needs, values, and capabilities, as suggested by the OECD Human-Centred Values and Fairness Principle: «AI should be developed in a way that is consistent with human-centred values, such as fundamental freedoms, equality, fairness, rule of law, social justice, data protection and privacy, and consumer rights and commercial fairness.» (OECD, 2021). The primary assumption of this study is to consider trustworthiness and human-centric principles as grounding elements of Responsibility in AI governance, assuming that responsible AI governance is the framework that allows respecting trustworthy AI principles.

Table 1 presents the requirements and related criteria in detail, linking the principles to the operational domain. In this

scheme, addressing information for all AI stakeholders designing, developing, deploying, implementing, using, or being affected by AI is compelling. The list includes (but is not limited to) companies, organisations, researchers, public services, government agencies, institutions, civil society organisations, individuals, workers and consumers.

Table 1: Details of requirements and criteria for Responsible Human-Centric AI Governance

Principle	Key Requirement	Criteria	Main AI-related threat
Respect for human autonomy	Human agency and oversight	Fundamental rights	Given the reach and capacity of AI systems, they can also negatively affect fundamental rights.
		Human agency	Users should be able to make informed autonomous decisions regarding AI systems.
		Human oversight	It helps ensure that an AI system does not undermine human autonomy or cause adverse effects.
Prevention of harm	Technical robustness and safety	Resilience to attack and security	AI systems [...] should be protected against vulnerabilities that can allow them to be exploited by adversaries.
		Fallback plan and general safety	AI systems should have safeguards that enable a fallback plan in case of problems.
		Accuracy	It pertains to an AI system’s ability to make correct judgments.
		Reliability and Reproducibility	It is critical that the results of AI systems are reproducible and reliable.
	Privacy and data governance	Privacy and data protection	AI systems must guarantee privacy and

			data protection throughout a system's entire lifecycle (GDPR).
		Quality and integrity of data	The quality of the data sets used is paramount to the performance of AI systems.
		Access to data	Only duly qualified personnel with the competence and need to access an individual's data should be allowed to do so
	Societal and environmental well-being	Sustainable and environmentally friendly AI	AI systems promise [...] must be ensured that this occurs in the most environmentally friendly way possible.
		Social impact	AI systems can contribute to the deterioration of social skills and affect people's physical and mental well-being.
Fairness		Society and Democracy	The use of AI systems should be carefully considered in situations involving the democratic process, including political decision-making and electoral contexts.
Explicability	Transparency	Traceability	It facilitates auditability as well as explainability
		Explainability	Technical explainability requires that the decisions made by an AI system can be understood and traced by human beings.
		Communication	Humans have the right to be informed that they are interacting with an AI system.
Fairness	Diversity, non-discrimination, and fairness	Avoidance of unfair bias	Data sets used by AI systems (both for training and operation) may suffer from the inclusion of

			inadvertent historical bias, incompleteness, and bad governance models.
		Accessibility and universal design	Systems should be user-centric and designed to allow all people to use AI products or services, regardless of their age, gender, abilities, or characteristics.
		Stakeholder Participation	It is advisable to consult stakeholders who may directly or indirectly be affected by the system throughout its life cycle.
	Accountability	Auditability	It entails enabling the assessment of algorithms, data, and design processes.
		Minimisation and reporting of negative impacts	The ability to report on actions or decisions that contribute to a certain system outcome and respond to the consequences of such an outcome must be ensured.
		Trade-offs	Tensions may arise between requirements, leading to inevitable trade-offs that should be addressed rationally and methodologically within the state of the art.
		Redress	Accessible mechanisms should be foreseen to ensure adequate redress.

Source: author's elaboration of the main AI-related threats based on information and classification provided in the "Ethics Guidelines for Trustworthy Artificial Intelligence by the High-Level Expert Group on AI", 2019.

2. Methodology of this study

After having broadened the concepts of responsible human-centric AI and trustworthiness, the study follows the logical steps:

- Build a detailed list of features and boundaries of responsible governance for the public sector. Building upon the requirements for RHAIG, the study evaluates current solutions and examples of public services and responsible governance made with and by AI. These paradigms contribute to building a knowledge base about what the model should answer AI's expected and potential role.
- Define a general-purpose qualitative model for RHAIG. The model embeds all the grounding elements of responsible and human-centric AI governance in the public sector. To this end, the author evaluates frameworks, methods, and concepts that can take responsibility and operational tasks by performing grey and academic literature analysis on existing AI governance frameworks. The academic survey considers the existing scientific articles published in Scopus and found through the query "*responsible + human + artificial intelligence + governance + framework.*"
- The grey analysis asks for public information from private organisations and agencies; the author used the same query to find reports, white papers, recommendations, and policy statements. Once selected for their relevance, all the documents are analysed to find differences and commonalities in the approach to AI governance. The RHAIG model includes the most relevant features of existing frameworks; all the elicited knowledge results in a general model (Figure 2) and requirement mapping (Table 2).

- Finally, a matching exercise helps start assigning a minimum set of skills and competencies to ensure the implementation of responsible human-centric AI deployments in the Public Sector. The author deploys a recent EU report on AI for Public Sector skills and procedural framework to evaluate which competencies are suitable for each responsibility and role's layers in RHAIG (Table 3).

3. The Responsible Appropriation of AI in the Public Sector

Adopting AI in the public sector forces EU public administrations to «face the challenges of implementing AI solutions» (Grimmelikhuijsen, S. Tangi, L., 2024). A recent EC Public Sector Tech Watch observatory report mapped over a thousand AI cases in Europe, and its findings evidence the difference in the adoption of AI between the national and local levels. It identified a substantial difference: the former case is more interested in «improving processes and bureaucratic tasks». At the same time, the latter mainly uses AI for «enhancing the implementation of public services» (European Commission, 2024b). Given this difference, the author opted for a general approach to avoid territorial levels and consider three aspects of the process: the challenges, the governance layers, and the responsibility assignment. Trustworthy AI systems are expected to contribute to the public good; in both local and national contexts, all the potential solutions to AI governance shall consider and coherently balance the effort to «prevent possible harm that AI solutions could have to the individual and the community / collective levels» (Andrus, M., & Villeneuve, S., 2022). Individual risks relate to Privacy, Bias, Identity Misrepresentation, and Data Misuse. Collective risks consider the following: Expanded Surveillance

(invasive for targeted categories), Misrepresentation of Categories (misalignment with identity or experience), and Private Control for Fairness goal (potential blind spots in unfairness). Mindful of the opportunities of AI in the Public Sector and leveraging the capability to «enhance or optimise existing public services and create new services» (Manzoni M., *et al.*, 2022), public administrations started asking themselves whether they could be held responsible for the adverse consequences AI systems might generate; the obligations of the state concerning AI adverse effects ascribe responsibilities to those who develop and implement AI systems. It turns out that policymakers face a dilemma: the obligation to protect citizens from potential algorithmic harms coexists with the mandate to increase efficiency and enhance the quality of digital services (Molinari F. *et al.*, 2021). Several ways exist to build coherent and solid control frameworks for Trustworthy human-centric AI governance; the most common leverage are Rights and liabilities (protection), Command and control (penalties), Administrative oversight (agencies), Incentives (tax credit), Market-harnessing controls (non-economic driven), Public infrastructure (informing values), Mandatory disclosures (performance-related metrics), Public compensation (revenues compensating for harms). In this context, policymakers have the opportunity and duty to guide the development of fair AI systems that benefit society. At this scope, the European AI Act, which formally entered into force on August 1st, 2024, aims to foster trustworthy AI in Europe by ensuring that «AI systems respect fundamental rights, safety, and ethical principles while addressing the risks of AI models» (European Commission, 2021).

3.1 Key challenges for AI governance for administration

The policymakers' challenge is twofold: to govern AI, algorithms, and related automated processes, and govern with and by AI, using algorithms and computerised methods and systems to enhance and improve public services. They constitute different perspectives brought by the pervasive character of AI-enabled processes in the public sector:

- Governing AI requires a framework or a set of rules and laws, such as data regulatory regimes and practices, to address the primary driver of AI development, the conjunction of a massive quantity of data with innovative machine learning algorithms.
- Governing with and by AI represents the effective use and value AI can offer governments when redesigning internal administrative processes to enhance the quality and impact of public services. It still requires humans to be “in the loop” by using, controlling, and supervising technology that reinforces public offerings' capacity. The implementation of AI-based public services, however, requires a deeper understanding of potential risks associated with the expected benefits.

The increased maturity of citizens' interaction with digital means helps the Public Sector provide accurate, efficient services. In this context, AI-based technologies shall be considered a component of complex socio-technical-economic systems, and they generate several challenges in identifying roles and responsibilities in service delivery and in finding balanced governance. Each approach to governance comes with benefits and downsides, so effective AI governance in the appropriation process should strategically combine them to maximise impact and minimise harm or risks. AI appropriation helps appreciate the number of different

stakeholders involved in the governance of AI-based solutions for the government. «By this term [we mean] the union of adoption and implementation, also bearing in mind that individual and group users of a certain technology after it is embedded in an organisational setup make changes to both the technology and the environment, which inevitably feed back into both steps of the singled-out process» (Molinari F. *et al.*, 2021). The first challenge becomes clear: AI intensifies inequality between people with access to its services and those without access and between those who benefit from it and those who do not, causing the “digital divide”. Combining different data sources and predictive tools generates new insights into citizens, leading to unequal rights among citizen groups based on their use of AI-based services. Other challenges due to the seamless improvement and development of new digital solutions are mentioned in (Council of Europe, 2019): the problems of ‘many hands’, ‘humans in the loop’, and the ‘unpredictable effects’ of complex dynamics. The ‘many hands’ problem arises when attempting to identify the responsibilities for errors and harms resulting from the development and operation of AI systems based on several critical components. The ‘Human-in-the-Loop’ problem arises when considering that many tasks previously performed by humans are now partially undertaken by machines without control. The “unpredictability” problem exists when trying to identify, anticipate, and prevent adverse events arising from the interactions of two or more AI-based systems or solutions (e.g., two algorithms). When considering the process of AI appropriation, these challenges extend to the number of different stakeholders and procedures, thus introducing systemic issues and additional governance challenges: Multi-level governance between national and local levels, Accountability of political actions and automated systems, *Limited skills* among public servants, and *Limited Economic capacity* of local authorities.

3.2 The layers of AI governance within the appropriation procedures

Public policymakers need to assume a front-and-centre role in ensuring beneficial AI systems and, simultaneously, «face the challenges and overcome critical factors» (Misuraca, G., Van Noordt, C. 2020). A layered model is helpful when considering how policies can govern AI, aiming to define appropriate behaviour for AI and autonomous systems. This study deepens the application of an existing four-layer model proposed by (Gasser, U., Almeida, V. A., 2017):

- Technical layer. It is the foundation of the AI governance ecosystem, including the algorithms and data from which it is built. It establishes who is responsible for norms, regulations, and legislation.
- Ethical layer. It articulates high-level ethical concerns that apply to all types of AI applications and systems, including human rights principles introduced by trustworthy AI. It establishes who is responsible for the criteria and principles.
- Social and Legal layer. It addresses the process of creating institutions and allocating responsibilities for regulating AI and autonomous systems. One starting point for specific norms aimed at regulating AI can be the principles and criteria that emerge from the ethical and technical layers, in addition to pre-existing and more general national and international legal frameworks. It established who is responsible for data governance, algorithmic accountability, standards, and the evaluation of trade-offs in the design of AI systems.
- Evaluation layer. It covers as much of the lifecycle as possible to ensure that the focus considers productive deployment and evaluation. It establishes who is responsible for ex-ante and

continuous ex-post impact assessment and cost-benefit analysis of the acquired AI-based service.

3.3 Theoretical models for allocating responsibilities

The proposed RHAIG framework includes *Responsibility* layers. The “responsibility models” are mandatory when approaching high-risk categories identified by the AI Act; nevertheless, they are instrumental in correctly identifying how to uniquely assign responsibilities and formalities when facing systems’ faults or direct negligence. A report from the Council of Europe (Yeung, K., 2019) outlines four broad heuristic responsibility models addressing stakeholders involved in AI appropriation:

- Intention/culpability-based models. They focus primarily on the voluntariness of the agent’s conduct. They can be interpreted as requiring the satisfaction of the ‘control’ condition, demonstrating that the agent was causally responsible for the legally proscribed conduct as the agent had a free and voluntary choice concerning whether to act, notwithstanding the harmful consequences of the agent’s conduct.
- Risk/negligence-based models. They refer to harm as a reasonably foreseeable consequence of the computational systems’ actions and decisions, and can be ascribed to the human developers of computational agents and systems.
- Strict Responsibility. Responsibility attaches to the agent without proof of fault, so legal responsibility for rights violations attaches to those who cause them, regardless of whether the responsible agent engaged in conduct that breached a legally specified standard of conduct.

- Mandatory insurance schemes. It could be established on a ‘no-fault’ basis by establishing an insurance fund to which all those harmed by the operation of these technologies could have recourse. It admits to simply requiring stakeholders in the AI appropriation value chain to take out mandatory liability insurance.

No one policy option is perfect; each has its drawbacks and should be viewed within the context of an entire toolbox from which policymakers can draw. The harms of AI can be varied, and so too should the policy instruments used to address them if an effective change is to be made. However, the concept of responsibilities is a common characteristic of all the possible models and governance frameworks.

4. A Qualitative Model for Responsible Human-Centric AI Governance

4.1 Map of existing frameworks

Few researchers focus on concrete frameworks that facilitate the adoption of artificial intelligence. Recent works have compared existing literature frameworks suitable for helping to adopt AI in the public sector (Pechtor, V., & Basl, J., 2022); (Zuiderwijk, A., *et al.*, 2021). The analysis discussed the weight that seven main requirements have in each proposed framework: 1-Coverage of the AI lifecycle; 2-Concrete steps and fields of action; 3-Granularity on multiple levels; 4-Agile aspects (e.g., MLOps); 5-Consider AI maturity; 6-Incorporation of other methods (e.g., Technology–Organisation–Environment Framework, Diffusion of Innovation); 7-Inclusion of ethic or human-centric elements. In

total, 11 relevant articles suggest frameworks covering different aspects of AI in public services. The variety ranges from the prevention of discrimination, assessing the impact of AI, or even multi-layer conceptual frameworks:

- A realist perspective on AI-era public management doi.org/10.1145/3325112.3325261.
- AI innovation for advancing public service: The case of China's first administrative approval bureau. doi.org/10.1145/3325112.3325243.
- AI-Enabled Innovation in the Public Sector: A Framework for Digital Governance and Resilience doi.org/10.1007/978-3-030-57599-1_9.
- Deep learning meets deep democracy: Deliberative governance and responsible innovation in artificial intelligence. doi.org/10.1017/beq.2021.42.
- Evaluating the impact of artificial intelligence technologies in public services: Towards an assessment framework doi.org/10.1145/3428502.3428504.
- Exploring the implementation of artificial intelligence in the public sector doi.org/10.37394/232010.2020.17.9.
- Garbage in, garbage out: The vicious cycle of AI-based discrimination in the public sector doi.org/10.1145/3325112.3328220.
- Governance of artificial intelligence: A risk and guideline-based integrative framework doi.org/10.1016/j.giq.2022.101685.
- IoT and AI for smart government: A research agenda doi.org/10.1016/j.giq.2019.02.003.
- OECD Recommendation of the Council on Artificial Intelligence OECD/LEGAL/0449.

- Value-Based Guiding Principles for Managing Cognitive Computing Systems in the Public Sector doi.org/10.1080/15309576.2021.1879883.

Existing academic frameworks for the public sector are relatively few and vary in focus and application. Those covering the AI adoption process are more conceptual, while frameworks that help identify concrete fields of action are still scarce. “AI maturity” is considered in only four out of 11 proposed frameworks. As a general consideration, finding commonalities among these proposals remains difficult at the current time.

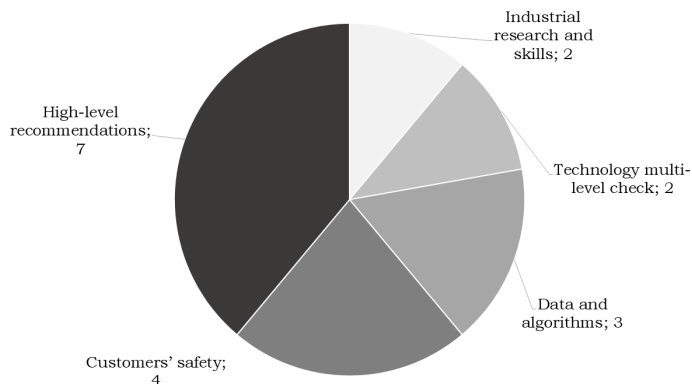
Analysing grey literature released by private and other initiatives yields usable results. The White Paper released by TNO (Veenstra, A.F. *et al.* 2021) considers 17 AI frameworks by focusing on a survey of current regulatory systems for AI and an analysis of existing ethical frameworks and methods. The author updated the list by substituting the revised OECD proposal and adding the work by Wirtz on the risks associated with public administrations. In total, the author considers 18 AI frameworks embedding ethical elements for the analysis:

- AI Guidelines (Deutsche Telekom)
- AI in the UK: ready, willing and able? – UK House of Lords, Select Committee on AI
- AI Policy Principles – Information Technology Industry Council (ITI)
- An Ethical Framework for a Good AI Society (Led by Prof. Luciano Floridi)
- Artificial Intelligence and Machine Learning: Policy Paper – Internet Society
- Automated and Connected Driving: Report – Federal Ministry of Transport and Digital Infrastructure, Ethics Commission

- Charlevoix Common Vision for the Future of Artificial Intelligence – Leaders of the G7
- Dynamic impact assessment methodology for responsible algorithmic decision-making (TNO)
- Ethically aligned design – Institute of Electrical and Electronics Engineering
- Everyday Ethics for Artificial Intelligence (IBM)
- How can humans keep the upper hand? Report on ethical matters raised by AI algorithms – French Data Protection Authority (CNIL)
- Montréal Declaration for Responsible Development of AI – Université de Montréal
- Preparing for the future of Artificial Intelligence – US National Science and Technology Council, Committee on Technology
- Principles for Accountable Algorithms and a Social Impact Statement for Algorithms – Fairness, Accountability and Transparency in Machine Learning
- The dark sides of Artificial Intelligence: An integrated AI governance framework for public administration, Wirtz (2020)
- Tools for trustworthy AI: A framework to compare implementation tools for trustworthy AI systems, OECD (2021).
- Top 10 Principles for Ethical Artificial Intelligence – UNI Global Union
- Z-Inspection initiative, led by Professor Roberto Zicari (Arcada University of Applied Sciences, Helsinki, Finland) and supported by several researchers worldwide.

Figure 1 summarises a brief author's at-purpose classification of the 18 frameworks. Most of them are recommendations (7), followed by narrower topics such as privacy and safety (4), data use (3), technology assessment (2), and industrial research or development (2).

Figure 1: Classification of existing responsible AI frameworks by considering their primary focus



Source: author's taxonomy and own elaboration based on desk research and analysis of listed frameworks.

The analysis of publicly available information on these frameworks revealed that they incorporate various ethical standards, but do not clarify who determines how these standards are applied; their optimal use is during an ex-post assessment of the impact of AI systems. Insufficient evidence supports the definition of a narrow implementation of responsible AI governance in the Public Sector. However, the analysis pointed out two relevant criteria to be of utmost importance for the definition of the author's proposed framework:

- A transdisciplinary approach that goes beyond involving public organisations, AI developers, and legal practitioners must incorporate the views of policy officials, regulatory bodies, and the public. All these stakeholders shall proactively ask themselves the above questions when using AI in algorithmic decision-making.
- There is a need to “Avoid the concentration of power” when distributing governance roles.

4.2 Definition of a multi-layered framework

When building the proposed framework, the author considers the complex dynamics ruling the cooperation among stakeholders and the flexibility of public-private AI appropriation schemes. Since building or refining a framework case by case is not possible or convenient, it became easier to adopt a qualitative model. The model aims to support the user, mainly the public authority or the policymaker, in detecting any activity that could lead to potential ethical warnings and assigning it to a procedural governance task. Moreover, it suggests focusing on those actions that can mitigate or eliminate the risk of ethical non-compliance in public processes and services when possible. Three aspects ground the choice for a qualitative typology:

- The ethics and responsibility principles are still at an abstraction level that makes it difficult, if not impossible, to measure their application or fulfilment.
- The fulfilment of ethics follows the progress of technology rather than measuring specific technological achievements or refinements.
- A qualitative model is more flexible than a quantitative one and better embeds the responsibility by design paradigm.

The approach of the author's study gathers these untold recommendations in the definition of a possible procedural model based on four fundamental pillars:

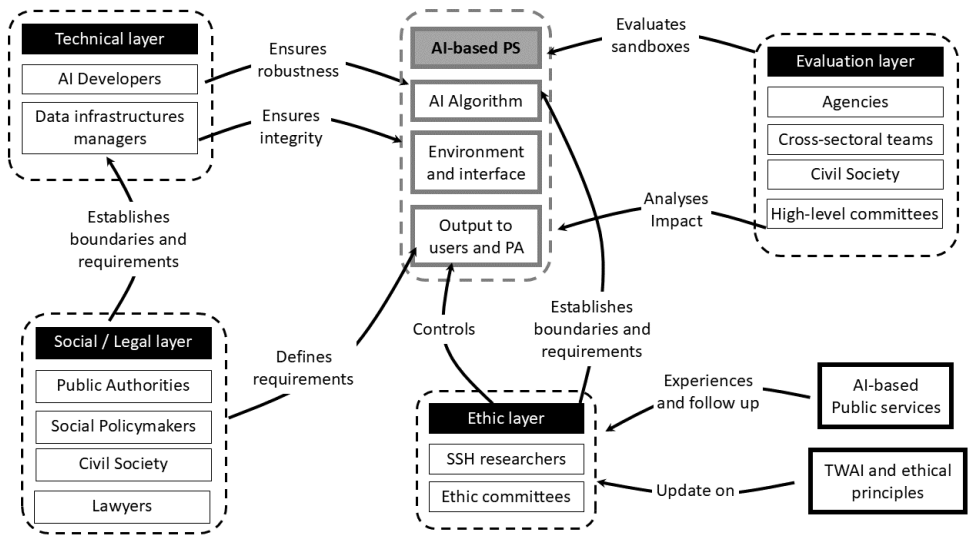
- Include a check on all criteria (Table 1) to evaluate any potential ethical breach.
- Consider the AI lifecycle as the sequential merging of three phases: AI design, AI appropriation, and AI (service) monitoring/evaluation.
- Require and check that all stakeholders are actively engaged

in the AI lifecycle, even if through different means.

- Identify rules and assign responsibilities to each stakeholder to ensure fair risk sharing and increase care for AI’s overall quality.

Figure 2 introduces the schematic responsible human-centric AI governance model graphical representation. Each dashed area (governance layer) includes responsible and operative (who acts) stakeholders, while the arrows establish the connection on responsibilities (who sets the rules). For the sake of readability, the figure shows only the most relevant interdependencies among stakeholders of the various governance layers and the AI-based public services.

Figure 2: Schematic governance model representation



Source: Author’s own elaboration, based on the match of Responsibility Models (Council of Europe, Yeung, K., 2019) and AI Governance Layers (Gasser, U., Almeida, V. A., 2017)

As said, responsible governance is the result of a refining exercise on each service and for each public authority providing the service. The customisation process could add or remove one or more stakeholders because of the technology's characteristics; accordingly, the scheme also includes double roles, especially in the evaluation layer. This layer oversees performing gap analysis in case some criteria are not adequately addressed and opts for a regulatory sandbox for the service's testing phase. On the other hand, the social and legal layers often decide which model of responsibility sharing to use, as they oversee the economic sustainability of the service and assign tasks to AI developers. The Ethics layer translates the key requirements into operational rules for the technical layer. Thanks to the experiential follow-up of other existing AI-based services, it provides enhanced control over the impact of the service on ethical principles. The approach of the governance model focuses on the assignment of clear roles and rules; therefore, when implementing the governance model, it is helpful to refer to a schematic procedure. In conclusion, the proposed governance model and the checklist method to implement it follow the strategies adopted by many of the other analysed frameworks. This study enhances this approach by adding three characteristics:

- The stakeholders' engagement ensures an adequate mapping of all the steps in the AI appropriation phase into the checklist with their responsibilities.
- The evaluation layer fulfils the need to constantly monitor the service's impact during its appropriation and operational phase.
- It embeds the "responsibility by design" paradigm, intending to support the public sector in operating ethically sustainable services from the very beginning and avoid ex-post legal monitoring by authorities.

Table 2 completes the governance model's scheme by regrouping all the missing details and specifying the tools and methods (third column). Although detailed, the table can be likewise refined. The nature of a governance tool enables service owners, providers, and public authorities or policymakers to fill all the table's cells with precise specifications and map roles for each engaged stakeholder.

Table 2 RHAIG framework detailed responsibility map

<i>Layer</i>	<i>Who “acts” and “is responsible for” in the layer</i>	<i>Who “sets the rules for” and “controls” the layer</i>	<i>With which “tools & methods”</i>
Technical	AI developers IT & Data Infrastructures managers IT and AI Research Centres	Public authorities National legal framework(s) Ethic committees (third parties)	Algorithm robustness check Ethic compliance (mainly security, bias, and exclusion)
Ethic	Ethic policy makers Social science researchers Ethic committees (third parties)	Ethic committees	the Implementation of TWAI principles Follow-up of AI services
Social / Legal	Public Authorities Social policy makers Civil society Lawyers	Public authorities Social policy makers	Definition of target territory and social boundaries Definition of models for allocation or responsibilities
Evaluation	Agencies (third parties) Cross-sectoral stakeholders' teams Civil Society High-level committees	Public authorities Legal framework	Constant KPI monitoring Sandboxes (and Living Labs) Citizens' feedback

Source: Author's own elaboration based on current research of mapped frameworks and responsibility models

The ex-post model's deployment highlights the RHAIG implementation in public AI-enabled services. Given the shortage of available data, the method relies on qualitative and Boolean analysis (YES / NO) to allow for a good evaluation rather than leveraging numerical indicators. From an operational perspective, the method requires checking which Trustworthy and Human Key Requirements are most relevant to the service and to what extent their governance has been fulfilled. The following checks are mandatory:

- Ensure to know (and have enough information about) which stakeholders are actively engaged in the whole AI-based service lifecycle: public authority adopting the service, AI developers, data infrastructure managers, social policy-makers, research centres, and – if possible – also some civil / citizen representatives.
- For each analysed criterion, all four layers of AI governance (technical, ethical, socio-legal, and evaluation) must be mapped as long as sufficient information about the existence and effectiveness of clear responsibilities among stakeholders for each layer is available.
- Discuss and analyse the possible consequences of the adopted AI based on all the criteria. The goal is to retrieve information on the Appropriate degree of human involvement, the risk of AI harm to individuals, the Plan for staff training (if necessary), and the Mechanisms for users' feedback.

5. A preliminary set of skills in Responsible Human-Centric AI Governance

Much work and research have been done to identify consolidated and emerging skills capable of AI implementation and

acquisition. Mindful of identifying the critical human-centric skills essential for working with and alongside AI, this study focuses on aligning the proposed framework model's needs with the existing peer-validated classification of mapped skills, rather than eliciting or defining new skills. Typical basic essential skills, such as Creativity, Critical Thinking, Collaboration, and Communication, can be developed «through education and training on the job» (Kumar, S., 2023). Human-centric AI, however, also requires understanding human behaviour and cognitive processes, human motivation, values, cognition, behaviour, learning, change, decision-making, and persuasion; human-centric AI design methods; Technology acceptance of interactive AI-infused systems; Design guidelines for Human-AI interaction; Algorithmic nudges and boosts, to name a few. The matching exercise goes in this direction by scouting additional skills through a cross-check between the skills/competencies classification and the requirements set by the proposed responsible AI framework, as mapped in Table 2.

The skills source for this study is a recent EC report, whose findings elaborate on the need to ensure the presence of proper competencies and governance practices to deploy AI solutions in the Public Sector. This approach facilitates the skills-matching exercise because it provides a global vision of the competencies needed to perform the governance practices effectively and assigns them to different responsibility and control layers, similar to our framework model approach. Notably, particular attention is given to structural governance practices, defined as «those that concern the identification of key decision-makers and their corresponding roles and responsibilities» (Medaglia, R., *et al.*, 2024), like the proposed framework. Moreover, its reliability leverages a synthesis of empirical research and grey and policy literature. The EC report defines three groups of

competencies: technical, managerial and policy, legal and ethical. For the sake of simplicity, the present study only considers the three groups; however, it is worth mentioning that the report also introduces three cross-cutting clusters drawing on the distinction between know-why, know-how and know-what types of knowledge. Matching the 56 different skills and competencies observed in the EC report with the requirements of the proposed framework brings the following as a starting base for the list of skills needed for ensuring the consistent implementation of responsible human-centric AI governance in the public sector:

1. Technical:

- Technical Design Thinking, *being able to approach innovation of AI technology in an iterative and user-centred way*
- Data quality assessment, *being able to assess the quality of data in AI contexts*
- Choice of machine learning techniques, *being able to know when to use a certain algorithm, tool or library in a specific situation*
- Compliance with AI standards, *being able to adhere to and develop AI based on ethical and legal technical standards*

2. Managerial:

- AI benefits understanding: *being able to understand the benefits of AI.*
- User centricity: *being able and willing to collaborate with users of AI digital services and valuing their feedback.*
- Risk anticipation and migration: *being able to anticipate and mitigate risks of AI (e.g. privacy, security and ethics).*
- Data-supported decision-making: *being able to make decisions based on data.*
- The choice to delegate to AI: *being able to consider relevant factors in deciding whether or not to delegate a public service or a process to AI.*

- Change management: *being able to manage changes in organisational processes introduced by AI.*

3. Policy-Legal-Ethical:

- Awareness of ethical implications: *being able to be aware of the implications of AI on ethical and moral issues.*
- Awareness of sustainability implications: *being able to be aware of the implications of AI on environmental sustainability*
- Policy Design thinking: *being able to approach AI policymaking in an iterative and user-centred way.*
- Collaboration with AI ethicists: *being able to judge when experts on AI ethics should be consulted.*
- Collaboration with domain experts: *being able to work together with domain experts from varying professional backgrounds.*
- Understanding legal and ethical frameworks: *being able to understand and be aware of relevant legal and ethical frameworks for AI.*
- Privacy and security literacy: *being able to understand and act on the issues, concerns and threats around privacy and security raised by AI.*

The list includes the minimum set of skills and competencies addressing responsible human-centric AI, whilst still not being exhaustive. Table 3 briefly addresses the skills mapping exercise by assigning the skills in the proposed scheme.

Table 3: Map of competencies and skills for the proposed RHAIG framework

Layer	Who “acts” and “is responsible for” in the layer	Who “sets the rules for” and “controls” the layer
Technical	Technical Design Thinking Data supported decision-making Collaboration with domain experts	Choice of machine learning techniques Collaboration with AI ethicists Change management
Ethic	Awareness of ethical implications Awareness of sustainability implications	Risk anticipation and migration
Social / Legal	Data quality assessment Understanding legal and ethical frameworks	Policy Design thinking Privacy and security literacy
Evaluation	Compliance with AI standards User centricity	Choice to delegate to AI AI benefits understanding

Source: Author’s own elaboration, based on the proposed framework and the analysis of “Competences and governance practices for artificial intelligence in the public sector”, 2024

Conclusion

This study provides a preliminary, methodologically grounded, albeit incomplete, answer to the central question of defining a model for Responsible Human-Centric AI Governance that incorporates ethics and methods defined by the European public sector’s trustworthiness policies. The proposed model provides added value to the mapped existing frameworks by analysing the ethical principles that consider the inner (disaggregated) layer of criteria. To the author’s knowledge, this detail is not considered now in the landscape of solutions for AI governance. It is noticed, however, that the OECD is currently developing a framework based on the Recommendation on AI (OECD, 2024) and the Tools for Trustworthy AI (OECD, 2021). The proposed structure is general-purpose and includes a checklist

covering various aspects, including the roles and duties of stakeholders. The model, therefore, requires information from multiple sources; extensive interviews or stakeholders' questionnaires optimise the quality, whereas desk-based exercises could limit its potential. The model's goal is to provide suggestions and recommendations on how risks of implementing uncompliant ethical AI in the Public Sector can be mitigated with procedures and by leveraging a minimum set of skills. The proposed model constitutes an attempt to guide public authorities willing to adopt AI-based solutions by suggesting how and with whom they engage and with which tools. Its procedures allow users to fill out the responsibility mapping table by spotting gaps identified by three parameters: WHAT (the criteria not adequately addressed and the possible action), WHO (the stakeholders mostly entitled to be responsible for that action), and the responsibility LAYER considered (if technical, ethical, legal, or evaluation). Another problem this study highlights is that, in the AI acquisition process, stakeholders often find ethical frameworks abstract and challenging to apply when developing or tuning systems, leaving unanswered questions, such as at what development stage an AI system should be tested against these principles. The proposed minimum set of skills serves as a starting point to address this knowledge and awareness gap. Yet, it is far from claiming to be a definitive solution for identifying the needs of AI practitioners in the context of human-centric AI. On the other hand, it shows that current mapped skills and competencies could play a relevant role in ensuring that ethics and trustworthiness are embedded in the human-centric AI deployment and acquisition by the Public Sector. Three main recommendations come from this study and represent interwoven issues:

- It is crucial to ensure the 'autonomy' of the responsibility assignment for any AI-based development. Establishing a

straightforward and flexible approach to jointly regulating AI seems relevant to ensure innovation while safeguarding fundamental values and preventing harm to people.

- Rules and regulations are useless without coherent training. One of the main challenges is motivating and engaging core stakeholders to ensure they act upon the given recommendations. Often, skills and knowledge constitute barriers and ensuring that all subjects are correctly trained on responsible procedures will smooth them out.

Investing in research on competencies for trustworthy, human-centric AI in the public sector should be a global policy. AI-based technologies continue to grow; generative AI applications are a prime example of how public servants require the latest competencies, rather than existing ones, to utilise large language model prompts for document drafting.

References

AI-HLEG (2019). High-Level Expert Group on Artificial Intelligence: Ethics guidelines for trustworthy AI.

Andrus, M., & Villeneuve, S. (2022, June). Demographic-reliant algorithmic fairness: Characterising the risks of demographic data collection in the pursuit of fairness. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 1709-1721.

Council of Europe (2019). Responsibility and AI: Council of Europe Study DGI(2019)05.

European Commission (2021). Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act).

European Commission (2024a). Public Sector Tech Watch: Mapping innovation in the EU public services: a collective effort in exploring the applications of artificial intelligence and blockchain in the public sector.

European Commission (2024b). Public Sector Tech Watch: Adoption of AI, Blockchain and other emerging technologies within the European public sector. Publications Office of the European Union, Luxembourg, 2024.

Gasser, U., Almeida, V. A. (2017). A layered model for AI governance. *IEEE Internet Computing*, 21(6), 58-62.

Grimmelikhuijsen, S. Tangi, L. (2024). *What factors influence perceived artificial intelligence adoption by public managers*. European Commission: Joint Research Centre. Luxembourg: Publications Office of the European Union. <https://data.europa.eu/doi/10.2760/0179285>, JRC138684.

Hak, A. (2022). Why AI governance is important for building more trustworthy, explainable AI. <https://thenextweb.com/news/ai-governance-critical-trustworthy-explainable-ai> (accessed on 18/10/2024).

Kumar, S. (2023). Developing Human Skills in the Era of Artificial Intelligence: Challenges and Opportunities for Education and Training. *Scholedge International Journal of Multidisciplinary &*

Allied Studies, 10(2), 11-19. <https://dx.doi.org/10.19085/sijmas100201>.

Manzoni, M., *et al.* (2022). AI Watch. Road to the adoption of Artificial Intelligence by the public sector.

Medaglia, R., *et al.* (2024). Competences and governance practices for artificial intelligence in the public sector, Luxembourg: Publications Office of the European Union. European Commission: Joint Research Centre, <https://data.europa.eu/doi/10.2760/7895569>, JRC138702.

Misuraca, G. and Van Noordt, C. (2020). AI Watch - Artificial Intelligence in public services, EUR 30255 EN. Luxembourg: Publications Office of the European Union. doi:10.2760/039619, JRC120399. <https://publications.jrc.ec.europa.eu/repository/handle/JRC126665>.

Molinari, F., *et al.* (2021). AI Watch. Beyond pilots: sustainable implementation of AI in public services. EUR 30868 EN, Luxembourg: Publications Office of the European Union. doi:10.2760/440212, JRC126665.

OECD (2021), Tools for trustworthy AI: A framework to compare implementation tools for trustworthy AI systems. *OECD Digital Economy Papers*, 312.

OECD (2024) Recommendation of the Council on Artificial Intelligence. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> (accessed on 18/10/2024).

Pechtor, V., Basl, J. (2022). Analysis of suitable frameworks for artificial intelligence adoption in the public sector. IDIMT-2022.

Tangi L., *et al.*, (2022), AI Watch. European landscape on the Use of Artificial Intelligence by the Public Sector, EUR 31088 EN, Luxembourg: Publications Office of the European Union. doi:10.2760/39336, JRC129301.

UN (2024). Resolution on the promotion of safe, secure and trustworthy artificial intelligence systems for sustainable development. A/78/L.49.

Veenstra, A.F. *et al.* (2021). AI: in search of the human dimension. <https://publications.tno.nl/publication/34638227/KTvX9Y/veens-tra-2021-ai.pdf>.

Yeung, K. (2019). *Responsibility and AI: A Study of the Implications of Advanced Digital Technologies (including AI Systems) for the Concept of Responsibility Within a Human Rights Framework*. <https://rm.coe.int/a-study-of-the-implications-of-advanced-digital-technologies-including/168096bdab>.

Zuiderwijk, A., *et al.* (2021). Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. *Government Information Quarterly*, 38(3).

6. AI AND DEMOCRACY: THE ROLE OF THE EUROPEAN PARLIAMENT IN SHAPING THE EU “AI ACT”

by Raffaella Cinquanta*

Abstract: Il saggio intende evidenziare il ruolo politico svolto dal Parlamento europeo nell’elaborazione del *AI Act* del 2024: l’inquadramento etico e antropocentrico, la focalizzazione sulla regolamentazione dell’impatto dell’IA sulla vita democratica e la salvaguardia della coerenza con i valori fondativi dell’UE. La ricerca segue il dossier legislativo, analizzando i rapporti delle commissioni speciale e interne del PE. L’obiettivo è definire, anche in un confronto con l’approccio del Consiglio, la visione politica alla base dell’azione normativa del PE: la necessità di ripensare il contratto sociale al centro delle democrazie definendo un nuovo modello di governance democratica edificata sulla tecnologia.

Parole chiave: Intelligenza artificiale, AI Act, Processo democratico, Tecnologia umano-centrica, Parlamento europeo.

Abstract: The paper aims at highlighting the political role played by the European Parliament in the drafting of the 2024 EU *AI Act*: placing it in an ethic and human-centred framework, focusing on regulating the impact of AI on democracy and guaranteeing its consistency with EU founding values. The research follows the “procedure file” process, analysing the reports of the EP special and internal committees. The goal

* PhD in History of Federalism of European integration, subject expert and lecturer at Insubria University, researching at Pavia University’s Historical Archive. raffaella.cinquanta@unipv.it.

Accettato Dicembre 2024 – Pubblicato Aprile 2025.

is to define, also in comparison with the Council's approach, the specific political vision which underpins the EP legislative action: the need to rethink the social contract at the heart of democracies, defining a new model of democratic governance built around technology.

Keywords: Artificial intelligence, AI Act, Democratic process, Human-centred technology, European Parliament.

Introduction

In the framework of the 2030 Digital Decade and in the belief that EU approach to Artificial Intelligence (AI) will define its future, in the first months of 2024 the European Union passed an *Artificial Intelligence Act (AI Act)*, following a Commission's proposal of the 21st of April 2021 and a political deal between the European Parliament (EP) and the Council, negotiated in December 2023. The main purpose of the regulation is to safeguard fundamental rights and democracy by setting up concrete rules and specific actions for the use of AI based on its potential risks and level of impact, thus promoting its excellence, transparency, and accountability. The aspiration is to set a global standard, establishing EU as a world-class center for a human-centric and trustworthy AI (Ulnicane, 2021; Cath *et al.*, 2018; Roberts *et al.*, 2021b; Roberts *et al.*, 2023). Regardless of its actual effectiveness in fulfilling its purpose (Laux *et al.*, 2022; Mügge, 2024), which is not under scrutiny herein, the *Act* is the result of a long and thoroughly debated legislative process, which ended on the 1st of August 2024 when it entered into force. The rationale of the debate has been to strengthen the EU's ability to compete globally by developing AI as a positive force for European society, and to build strategic leadership in high-impact areas such as environment, health, robotics, public sector, home affairs, transport, and

agriculture. However, applying AI to these sectors requires that its use be trustworthy and safe. Hence, the focus on AI-related ethics and risks at different levels and the need to address them legally through an array of complementary, proportionate, and flexible rules, guiding AI developers, operators, and users in those cases not covered by the ongoing national and EU legislations¹. The paper aims at highlighting the role played by the EP in the legal procedure that led to the final approval of the *AI Act*, which arguably consists in underpinning the use of AI to the fundamental values at the basis of European society and in installing its regulatory framework in an ethic and human-centred background, focused on safeguarding democracy and equality. In this regard, it appears worthwhile to recall the connection to the proposals by the Conference on the Future of Europe, the citizen-led series of debates held from April 2021 to May 2022 to share and shape EU's future².

1. The ethical issue: preliminary debate

The need to start an inclusive discussion specifically focused on AI ethics manifests from the very background of the long legislative process, which has led to the *AI Act*³. Following the Declaration of Cooperation signed by 24 EU Member States and

¹ For an overview of the Commission's approach and action see <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence#:~:text=The%20European%20approach%20to%20AI,people's%20safety%20and%20fundamental%20rights>.

² On the Conference see the archived webpage <https://wayback.archive-it.org/12/090/20230417172132/https://futureu.europa.eu/en>.

³ The Commission's initiative found its background in the conclusions of the European Council of October 2017 on Digital Europe and in the Joint Declaration on the EU's legislative priorities for 2018-2019 calling for the implementation of a Digital Single Market.

Norway on 10 April 2018, expressing their will to join in a common European effort to deal with AI-related issues, on the 25th, the Commission presented the Communication *Artificial Intelligence for Europe*. The aim was to activate a coordinated endeavour to make the most of AI potential benefits while tackling its potential risks, channelling AI-related economic, social and ethical issues in a three-pronged plan: boost public and private investment, arrange socio-economic changes⁴ and secure an appropriate ethical and legal framework to tackle AI-related liability or potentially biased decision-making issues. Given that, as in the fields of consumer protection and of the Digital Single Market, «EU's sustainable approach to technologies creates a competitive edge, by embracing change on the basis of the Union's values», the Commission meant to propose, by the end of 2018, ethical guidelines for AI addressing issues such as fairness, safety, security, social inclusion and algorithmic transparency with a focus on AI impact on fundamental rights, privacy, dignity, consumer protection, non-discrimination and data protection, based on the Charter of Fundamental Rights. The use of AI, especially when autonomous decision-making is involved, demands a pondering over the suitability of established rules and civil law on safety and, significantly, liability. In this respect, thus, «the EU can make a difference – and be the champion of an approach to AI that benefits people and society as a whole» (EC, 2018a).

In June, the Commission appointed 52 specialists to the High-Level Expert Group on Artificial Intelligence (AI-HLEG), gathering representatives of academia, business, and civil society

⁴ In the awareness that AI would transform deeply the job world, the Commission recommended Member States to modernise their education, training and labour systems accordingly and on the basis of the European Pillar of Social Rights. The Commission meant to contribute by financing AI-dedicated training courses through the European Social Fund and EU's following multiannual financial framework (2021-2027).

to support the implementation of this plan in two respects⁵. Firstly, the Group would provide recommendations on how to tackle AI-related mid- and long-term challenges feeding into the legislative process to define a next-generation digital strategy, while drafting the above-mentioned ethics guidelines in accordance with the activity of the European Group on Ethics in Science and New Technologies and of the European Union Agency for Fundamental Rights. Secondly, the Group would help the Commission gather a wide community of stakeholders in a “European AI Alliance” to establish an open policy dialogue on AI. The Alliance was launched in June 2018 and regularly engaged approximately 6000 stakeholders in events, public consultations, and online forums⁶.

Shortly after, in December 2018, the Commission tried to involve States in the matter permanently, through a *Coordinated Plan*, to be updated yearly, proposing synergies and joint actions on AI with Member States, Norway and Switzerland in areas of public interest (i.e. healthcare, transport and mobility, security and energy). The goals of the initiative were to maximize investments through partnerships⁷, create large and secure common European data spaces in compliance with the General

⁵ On the AI-HLEG see <https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai>.

⁶ Four European AI Assemblies took place in June 2019, October 2020, September 2021 and November 2023. The Alliance has been operating also after the AI-HLEG mandate closed, sharing best practices and helping AI developers and stakeholders to self-assess their AI systems through the Assessment List for Trustworthy Artificial Intelligence (ALTAI). On the Alliance see <https://digital-strategy.ec.europa.eu/en/policies/european-ai-alliance>.

⁷ Since investment levels for AI in the EU were low and fragmented, the plan foresaw higher synergies to raise at least €20 billion of public and private investments in research and innovation till the end of 2020 and more than €20 billion per year in the following decade. Joint actions would include national AI strategies by mid-2019, a new European AI public-private research and innovation partnership, an AI scale-up fund to support startups and innovators, development and connection of EU world-leading centres for AI (through Digital Innovation Hubs and the launch of a European Innovation Council).

Data Protection Regulation, foster talent, digital skills and life-long learning with a focus on human-centred AI⁸, and build an ethical and trustworthy AI, respecting fundamental rights, to be replicated at global level. A new Commission service, the AI Watch, would help monitor the development of AI in the EU (Van Noordt and Misuraca, 2022) and the implementation of the Coordinated Plan⁹. On 8 April 2019, the AI-HLEG delivered the *Ethics Guidelines for Trustworthy AI*¹⁰ endorsed by the Commission in the Communication *Building trust in human-centric artificial intelligence*. The guidelines established seven key requirements for an AI to be trustworthy: human agency and oversight; technical resiliency and safety; privacy and data governance; transparency; diversity, non-discrimination, and fairness; societal and environmental well-being; accountability. After a piloting phase involving stakeholders, in July 2020, these requirements were translated into the *Assessment List for Trustworthy AI*, a concrete tool to guide developers and deployers in ensuring a risk-free AI. Besides, on the occasion of the first “European AI Alliance” Assembly on 26 June 2019, the AI-HLEG presented the document *Policy and investment recommendations of AI*. In the belief that EU and its Member States were placed in an exceptional position to seize the benefits of a minimum-risk AI, the document provided 33 recommendations to steer trustworthy AI towards sustainability, competitiveness and inclusion to empower and protect human beings, especially by focusing on enhancing data and infrastructure, skills and education, governance and regulation,

⁸ In this regard it is worth mentioning that the plan highlighted the importance to include educational programmes of disciplines not directly related to AI, such as law.

⁹ See https://ai-watch.ec.europa.eu/index_en.

¹⁰ The document followed the publication of the first draft in December 2018, which was debated in an open consultation and in meetings with representatives from Member States.

funding and investment in the areas of society at large, private and public sectors, research and academia. Finally, building on the work of the AI-HLEG, on 19 February 2020, the Commission released a *White paper on AI: a European approach to excellence and trust*¹¹, built around two main subjects. Firstly, it dealt with the policy framework to align measures at European, national, and regional levels, based on the partnership between private and public sectors, in order to create an “ecosystem of excellence” starting in research and innovation and accelerating the adoption of AI-based solutions also by small and medium-sized enterprises. Secondly, the paper sets the key elements of a regulatory framework for AI to create an “ecosystem of trust”, conceived as a policy objective in itself, through the compliance with EU rules and a human-centric approach respectful of EU values and citizens’ rights¹². Therefore, in the preliminary discussion launched by the Commission, it would seem that the ethical and human-centred matter, dealing mainly with the privacy issue, was debated on an equal footing with the economic and competitiveness goals.

2. The impact of AI on democracy: the role of the European Parliament

It can be argued that the EP would take responsibility for granting complete prominence to the former as a precondition for

¹¹ The *White Paper* was accompanied by a *Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics*, and by two topic-related communications: *Shaping Europe’s digital future* and *A European Strategy for Data*.

¹² The *White Paper* was discussed in a public consultation till June, receiving over 1215 contributions (including 400 position papers) through the online tools of the European AI Alliance (see <https://digital-strategy.ec.europa.eu/en/library/white-paper-artificial-intelligence-public-consultation-towards-european-approach-excellence-and>).

the latter. And that in doing so, it injected the AI ethical issue with a more ideal trait – envisaging a kind of technology serving humanity and pursuing the common good – and with a more long-term political approach, focused on AI impact on democratic life and military purposes (Nardocci, 2024; Roberts *et al.*, 2021a). In September 2020, the 33 members of the Special Committee on AI in a Digital Age (AIDA) of the EP started their work, which lasted till March 2022¹³. The EP entered the AI-policy debate stressing the need for Europe to speak with one voice through a «human-centric» and «holistic approach for a common, long-term position that highlights the EU's key values and objectives relating to AI in the digital age that ensures that the digital transition is (...) consistent with the Charter of Fundamental Rights» (EP, Special Committee on Artificial Intelligence in the Digital Age, 2022b). The European legislature had been dealing with AI- and digital-related democratic issue since 2017, with a series of resolutions on civil law, civil liability, intellectual property and ethical aspects of AI, robotics and related technologies, and, most significantly, approved guidelines on the use of AI for military purposes and in the health and justice sectors. The AIDA was appointed in June with the mandate to assess the impact and challenges of AI on the EU economy and society and present the EP with a strategic plan defining EU medium- and long-term objectives¹⁴. As a part of its

¹³ See <https://www.europarl.europa.eu/committees/en/archives/9/aida/home/welcome-words> and the *Activity Report following AIDA Mandate 2020-2022*.

¹⁴ In 28 meetings and 11 public hearings the AIDA, chaired by the Romanian Dragoş Tudorache of the Renew Europe Group, dealt with the following topics: AI in health and combatting pandemics; AI and the Green Deal; AI Diplomacy and Governance in a Global Setting; AI, Cybersecurity and Defence; AI and Competitiveness; AI and the Future of Democracy; AI and the Labour Market; AI in agriculture and food security; AI and financial services; AI and the Data Strategy; AI and Transport; and lastly, AI and Bias. The issues were debated with a broad range of AI experts and stakeholders and the results were published in a series of AIDA working papers. In addition, AIDA met several times with the Commission's Executive Vice-President Margrethe Vestager and the

work, the AIDA convened an Interparliamentary Committee meeting on 8 November 2021 with national parliaments, in order to discuss the future regulatory models, exchange information, and best practices¹⁵. The AIDA report, presented on 5 April 2022 by German rapporteur Axel Voss (Group of the European People's Party)¹⁶ and considered an «urgent call to action», was debated by the EP plenary on 3 May 2022 and adopted as a Resolution on AI in the digital age¹⁷. As for the impact and challenges of AI, AIDA had balanced its benefits towards risks in six case studies: health, the Green Deal, external policy and the security dimension, competitiveness, labour market, and the future of democracy. Thus, it was showed that EU still didn't meet any of the preconditions that «enable innovation to fully capture the potential of AI and other emerging technologies» due to a series of deficiencies in terms of digital single market: access and sharing of high-quality data, harmonized rules and standards, high regulatory burden, funding, research, skills and infrastructure (EP, Special Committee on Artificial Intelligence in the Digital Age, 2022b). In tackling this problem, the EP Resolution is consequential under more than one respect. First of all, it cautioned that if the EU does not «act swiftly and courageously, it will end up having to follow rules and standards set by others», often by non-democratic actors. Therefore, the Resolution called for a strong political commitment and a «more forward-leaning

Commissioner Thierry Breton, the French Council Presidency, the German Bundestag Study Commission on Artificial Intelligence. It also organized thematic workshops with the U.S. National Security Commission on Artificial Intelligence, the OECD, and experts on AI in various areas of application.

¹⁵ See: <https://www.europarl.europa.eu/relnatparl/en/aida-icm-on-artificial-intelligence-and-/products-details/20211007CPU37281>.

¹⁶The report was adopted by the AIDA on 22 March 2022, with 25 votes in favour, 2 against and 6 abstentions. Procedure file at [https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?lang=en&reference=2020/2266\(INI\)](https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?lang=en&reference=2020/2266(INI))

¹⁷ The Resolution was adopted by 495 votes to 34, with 102 abstentions.

mindset», still lacking at that time, to create a «European way» in the digitalised world, based on a human-centric approach, ethical principles, the concept of the social market economy and aiming for «AI to effectively become an instrument that serves people and society and pursues the common good and general interest». Thus, it presented a roadmap engaging the Commission, Member States and the EP itself to elaborate a clear legal framework in the form of regulations that should be «flexible, principle-based, technology-neutral, future-proof and proportionate (...) to the type of individual and/or societal risk incurred by the use of an AI system». In this regard, the EP indeed underlined the distinction, which will be acknowledged by the *AI Act*, between “high-risk” AI (requiring additional legislative strict safeguards) and “low-risk” AI (requiring transparency for users). As for the ethical and legal questions, the Resolution emphasized, more than any other institutional document released up till then, AI threatens fundamental rights, in particular from military developments such as lethal autonomous weapons systems, and democracy-detrimental applications such as mass surveillance and citizens profiling. Conversely, it suggested to prioritise policy options which, if supported by a proper infrastructure, education, and training, could unlock AI’s potential in the areas of health, environment, climate change, capital and labour productivity, sustainable growth, and job creation. Finally, in line with the Commission’s *White Paper* but with a major focus on EU citizens’ awareness and responsibility, the report underlined the necessity for public services to achieve an “ecosystem of excellence”, providing citizens with digital and AI skills at all stages of education and employment, and “ecosystem of trust”, with awareness-raising campaigns aiming at promoting AI democratisation and by using AI in e-governance and e-health (EP, 2022).

3. Council and Parliament: two different visions

Sometime before the approval of the EP Resolution, namely on 21 April 2021, the Commission had released the Communication *Fostering a European approach to AI*, including a review of the 2018 Coordinated Plan and summarizing all inputs and recommendations received so far: the attention to the double-faced main AI feature (opportunities and risks), the focus on an European way to trustworthiness and the subsequent perspective of a global leadership in a human-centred AI development. Simultaneously, the legislative process was officially started through the *Proposal for a Regulation laying down harmonised rules on artificial intelligence*, the future *AI Act*¹⁸, as a part of a broader interrelated legal initiative for a trustworthy AI, including an AI Liability Directive¹⁹ and a revision of sectoral safety legislation²⁰. In particular, the *AI Act* dealt with prohibited AI practices, high-risk AI systems, transparency obligations, measures supporting innovation, governance, implementation, and codes of conduct. Still, economic- and ethics-related provisions appear to be considered at the same level of priority and importance. As far as the Council is concerned, between November 2021 and September 2022, the Presidency proposed a series of compromised texts which would merge on 6 December 2022 in a common position, a “general approach” adopted as the basis for the negotiations with the EP under the ordinary legislative

¹⁸ The proposal was discussed in September by the European AI Alliance (see <https://futurium.ec.europa.eu/en/european-ai-alliance/document/event-report-high-level-conference-ai-ambition-action>).

¹⁹ The proposal was advanced on 28 September 2022, with the goal to equip the internal market with rules on non-contractual civil liability regarding AI systems. It is relevant to underline that the proposal followed a legislative own-initiative resolution adopted by the EP in October 2020.

²⁰ The *Proposal for a Regulation on general product safety* was advanced on 30 June 2021, subject to a public consultation and adopted on 10 May 2023.

procedure²¹. Compared to the EP's democratic and ideal approach, the most relevant amendments by the Council can be traced in the definitions and the scopes of the regulation. Namely, it narrowed down the definition of AI to those systems developed through machine learning and knowledge-based approaches. It added an «horizontal layer» in the high-risk AI classification, to «ensure that AI systems that are not likely to cause serious fundamental rights violations or other significant risks are not captured», while adjusting the requirements so that they «are more technically feasible and less burdensome for stakeholders to comply with». Most significantly, an «explicit reference» was made to exclude “national security, defence, and military purposes from the scope of the AI Act», with several changes to provisions relating to the use of AI for law enforcement purposes (Council of the European Union, 2022). Therefore, the Council's position is rather plainly based on the political will to defend national sovereignty (symbolised by the security, defence, and military exemption) and economic interests (noticeable in the loosening of the requirements for stakeholders and enterprises). As far as the EP is concerned, discussions were led under a joint committee procedure by the Committee on Internal Market and Consumer Protection (rapporteur: Brando Benifei, Group of the Progressive Alliance of Socialists and Democrats, Italy) and the Committee on Civil Liberties, Justice and Home Affairs (rapporteur: Dragoş Tudorache, Renew Europe Group, chair of the AIDA, Romania)²². As proof of the extent of the impact of AI, the following Committees were also

²¹ As for the advisory institutions involved in the legislative process, the European Economic and Social Committee gave its opinion on 22 September 2021, the European Committee of the Regions on 2 December 2021, the European Central Bank on 29 December 2021. The European Data Protection Board and the European Data Protection Supervisor issued a joint opinion on 18 June 2021.

²² See the procedure file at [https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2021/0106\(COD\)&l=en](https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2021/0106(COD)&l=en).

associated with the legislative work with shared and/or exclusive competences: Environment, Public Health and Food Safety, Industry, Research and Energy, Transport and Tourism, Culture and Education, Legal Affairs²³. The Report drafted by the two co-rapporteurs was adopted by the Joint Committee on 22 May 2023 and, consistently with the AIDA Resolution, displays a clear, democracy-oriented political approach. As specified in the explanatory statement, the goal of the *AI Act* should be «to ensure both the protection of health, safety, fundamental rights, and Union values and, at the same time, the uptake of AI throughout the Union», a more integrated digital single market, and a legislative environment suited for entrepreneurship and innovation. This explains why, contrary to the Council, the EP Committees believed that «no AI system should be excluded ex-ante, either from the definition of artificial intelligence or by carving out exceptions for particular types of AI systems» and added to the list of forbidden practices the «predictive policing (...) in breach of the key principle of presumption of innocence». It also explains why the list of high-risk AI was expanded with systems who «have the potential, by influencing a large number of citizens of the Union, to impact the very functioning of our democracy», such as those used by candidates or parties to influence and count votes in local, national, or European elections, and with potentially-deceptive systems like «deepfakes impersonating real persons and editorial content written by AI». Besides, the report emphasised the responsibilities of AI users, urging them to play a more active role «in protecting the health, safety, and fundamental rights of EU citizens and EU values», by appointing «competent persons responsible for the human oversight of high-risk AI». As far as public authorities were concerned, the report stressed the

²³ Opinions in https://www.europarl.europa.eu/doceo/document/A-9-2023-0188_EN.html.

need to increase «democratic oversight, public scrutiny, and accountability, alongside more transparency towards the public on the use of AI systems in sensitive areas impacting upon people's lives». The political approach, calling for institutional accountability, is noticeable also in the provision regarding the governance and enforcement of the regulation. On one side, the role of the AI Board, the advisory body composed of representatives from each Member State, was enhanced in ensuring the uniform application of the Regulation, in providing recommendations to the Commission, and in acting as a forum for national supervisory authorities. On the other hand, a new enforcement mechanism by the Commission was proposed, to be triggered in cases amounting to infringements of at least three Member States, modelled on the Market Surveillance and Compliance Regulation. Furthermore, the Committees stressed the need to strengthen the involvement of stakeholders and civil society organizations in the provisions of the Regulation and, to «ensure that individuals are properly empowered when the use of an AI system infringes on their rights, but also in order to contribute to building trust in AI systems and their widespread use», added a dedicated chapter on «remedies for both natural and legal persons» (EP, Committee on the Internal Market and Consumer Protection, Committee on Civil Liberties, Justice and Home Affairs, 2023). On 14 June 2023 the EP, ahead of talks with EU Member States on the final shape of the Regulation, amended the *AI Act* in the view that its purpose should be «to promote the uptake of human centric and trustworthy artificial intelligence and to ensure a high level of protection of health, safety, fundamental rights, democracy and rule of law and the environment from harmful effects of artificial intelligence systems in the Union while supporting innovation and improving the functioning of the internal market»²⁴. To this end, the EP

²⁴ The amendments were approved by 499 votes to 28 with 93 abstentions.

proposed amendments aiming at ensuring that all operators covered by the Regulation should develop and use AI following the general principles defined in the AI-HLEG *Ethics Guidelines*. Under this respect, quite relevant is the amendment regarding the promotion of «a sufficient level» of AI literacy, referring to skills, knowledge and understanding that allows providers, users and affected persons «to make an informed deployment of AI systems, as well as to gain awareness about the opportunities and risks of AI and possible harm it can cause and thereby promote its democratic control». As for the prohibition of AI practices, the EP's effort to protect democratic rights is blatant. The list was indeed expanded to include intrusive and discriminatory uses of AI such as: systems that use subliminal techniques or deliberately manipulative or deceptive techniques; systems that exploit the possible vulnerabilities of a given person or group of persons with the aim of substantially distorting their behaviour by «appreciably impairing the person's ability to make an informed decision»; biometric categorisation systems based on sensitive or protected attributes; “real-time”/“post” remote biometric identification systems (except for the prosecution of serious crimes and only after judicial authorisation); systems used for social rating or predictive policing, emotion recognition systems in law enforcement, border management, the workplace, and educational institutions; creation of facial recognition databases violating human rights and the right to privacy. Particularly significant is the following amendment, which will be transposed in the *AI Act*: «The use of AI systems in migration, asylum and border control management should in no circumstances be used by Member States or Union institutions, agencies or bodies as a means to circumvent their international obligations (...) nor (...) infringe on the principle of non-refoulement, or deny safe and effective legal

avenues into the territory of the Union, including the right to international protection».

As for high-risk AI, the amendments focused on equality and non-discrimination. The list was expanded with systems intended to be used: as security components in the management of critical digital infrastructures; to assess the appropriate level of education of an individual «materially» influencing «the level of education and vocational training from which that individual will benefit or to which he or she will have access»; to «monitor and detect prohibited behaviour in students during tests»; to make or substantially influence decisions «on the eligibility of natural persons for health and life insurance»; to «evaluate and classify emergency calls from individuals» and, most of all, «to influence the outcome of an election or referendum or the voting behaviour of natural persons in the exercise of their vote». The EP also introduced obligations (transparency requirements and safeguards against illegal content) for general purpose AI (GPAI), which should always disclose its AI-generated content. From an institutional point of view, the EP proposed to establish an AI Office with legal personality and acting «in full independence», cooperating with Member States, national supervisory authorities, the Commission, and other EU institutions on the implementation and application of the Act. It also introduced the right to complain about AI systems with a national supervisory authority in case of regulation infringement (EP, 2023).

4. AI and the Conference on the Future of Europe

To better outline the political vision which underpinned the EP legislative action it is significant to point out a possible democratic influence that pushed the Assembly to face the issue

on the assumption that the AI-related challenges were forcing institutions and citizens to rethink the social contract at the heart of democracies, and therefore to define a new model of democratic governance built around technology. Besides the studies on AI provided by the European Parliament Research Service (EPRS), overall focusing on AI economic and social impact and, more importantly, on democracy-related issue²⁵, it appears of particular relevance the peculiar experiment in supranational participatory democracy that was the Conference in the Future of Europe (Moccia, 2021; Cinquanta 2022). Held from April 2021 to May 2022, its final report lists a series of proposals to reform the EU according to citizens' vision of its future. Significantly, the topic of AI was discussed in the Citizens' Panel dedicated to "European democracy/Values and rights, rule of law, security". On one hand, it was regarded as a security question: citizens recommended that EU institutions should «play a stronger role» and «protect individuals, organizations and institutions against new threats coming from cybersecurity breaches and the use of AI for criminal purposes. (...) because these threats are a serious national and European security concern. We recommend this because Europe should be a true innovator in this field». On the other hand, a beneficial use of AI was related to identity and educational developments: «We recommend that existing and emerging translation technologies such as artificial intelligence are further developed, improved and made more accessible so as to reduce language barriers and strengthen common identity and democracy in the European Union».

In the final report, the Conference stated that Europe should become a «world leader and standard setter in digital transformation and charter a European way to build an ethical, human-centred, transparent and safe digital society». Therefore,

²⁵ See the website <https://epthinktank.eu/?s=artificial+intelligence>.

Europe needs «to be ambitious in its approach» and fully use the opportunities offered by digitalisation, while at the same time managing its risks and challenges through «a possible future Charter of Digital Rights». As for the specific impact of AI on society and economy, the Conference prompted Europe to become a world leader in «guaranteeing human oversight of all processes involving artificial intelligence in work environments while fully utilizing the potential of trustworthy and responsible use of artificial intelligence, setting safeguards and standards that ensure transparency, interoperability, generate trust, enhance ease of use and avoiding any discriminatory or biased algorithms» (Conference on the future of Europe, 2022).

5. The negotiations and the approval of the AI Act

After the EP amendments, the interinstitutional negotiations began. The first four political trilogues, held on the same 14 June 2023, then on 18 July 2023, 2-3 October 2023 and 24 October 2023, agreed on the “less controversial parts” of the Commission’s proposal and found a compromise on the mechanism for classification of high-risk AI systems. In the fifth and final trilogue, held on 6-8 December 2023, the co-legislators closed the negotiations with a compromise on the most controversial political issues, such as the regulation of GPAI models and systems, their governance, as well as the prohibitions, especially the much-debated law enforcement block. Meanwhile, in compliance to the forthcoming regulations, in January 2024, the Commission launched a package of measures to support European

startups and SMEs in the development of trustworthy AI²⁶. It established the European AI Office as a part of the administrative structure of the DG for Communication Networks, Content and Technology²⁷.

As for the most politically marked aspects of the EP amendments, the compromise text seems to take some step back, namely – and quite significantly – in the matter of national sovereignty and interest. Firstly, the text did include the statement that, besides improving the functioning of the internal market and supporting innovation, the goal of the *AI Act* is «to ensure a high level of protection of health, safety and fundamental rights enshrined in the Charter, which includes democracy, rule of law and environmental protection». But all subsequent references to the risks addressed by the Regulation included only risks to health, safety, and fundamental rights, in accordance with the Council's mandate. Secondly, and most noticeably, the compromise text made it clear that the Regulation excluded national security. As for the forbidden AI practices, the compromised list included the prohibition of real-time biometric identification by law enforcement authorities in publicly accessible spaces, «but with some notable and clearly defined exceptions (...) now subject to a range of safeguards, including monitoring and oversight measures and limited reporting obligations at EU level». Other prohibitions were accepted in a limited form: untargeted scraping of facial images for the purpose of a creating or expanding facial recognition databases, emotion recognition but only at the workplace and in educational institutions (and with exceptions for safety and medical reasons), biometric categorization based on

²⁶ All documents available at <https://digital-strategy.ec.europa.eu/en/news/commission-launches-ai-innovation-package-support-artificial-intelligence-startups-and-smes>.

²⁷ See <https://digital-strategy.ec.europa.eu/en/policies/ai-office>.

certain specific beliefs or characteristics²⁸, and, significantly, predictive policing which assess or predict the risk to commit a criminal offence based solely on the profiling of a natural person or on assessing their personality traits and characteristics. Besides, the compromise agreement preserved all the exceptions for law enforcement authorities added by Council (i.e. derogation from conformity assessment²⁹, possibility to start testing of high-risk AI systems without prior authorisation), while the obligation for some deployers to conduct a fundamental rights impact assessment was lightened in order to make it easier to comply with. As for GPAI, defined as «AI systems based on GPAI models that can serve a variety of purposes», the compromise introduced horizontal obligations, such as updating and making available technical documentation to the AI Office and national competent authorities, and some additional requirements for models with systemic risks (model evaluation, risk assessments, risk mitigation measures, cybersecurity protection, reporting serious incidents to the AI Office and national competent authorities) to be achieved through codes of practice developed by the industry with the participation of the Member States, through the AI Board. As for governance, the compromise stated that while the market surveillance system on the

²⁸ As for real-time remote biometric identification systems in publicly accessible spaces for the purposes of law enforcement, the compromise text reinserts and increases the exemptions proposed by the Commission: their use is prohibited «unless and in as far as such use is strictly necessary for one of the following objectives: (i) the targeted search for specific victims of abduction, trafficking in human beings or sexual exploitation of human beings, as well as searching for missing persons; (ii) the prevention of a genuine threat of a terrorist attack; (iii) the localisation or identification of a person suspected of having committed a criminal offence, for the purposes of conducting a criminal investigation, prosecution or executing a criminal penalty for offences, (...) punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least four years» (Council of the European Union, 2024).

²⁹ However, the Council pointed out that «to secure this concession, it has been necessary to re-introduce in the text the wording regarding the control of the authorisation process by the Commission» (Council of the European Union, 2024).

national level would apply to AI systems, the new rules for GPAI provided for a more centralised oversight and enforcement. Thus, the list of tasks of the AI Board was extended to «give the Member States a stronger coordination role». Besides, two new advisory bodies were established: a scientific panel of independent experts for technical advice to the AI Office and market surveillance authorities (Member States could call upon experts to support their market surveillance activities) and an advisory forum made up of a balanced selection of stakeholders to provide input to the Commission and to the AI Board. Member States were also allowed to appoint at least one notifying authority and at least one market surveillance authority as national competent authorities (Council of the European Union, 2024)³⁰. Once the agreement was reached, the final legislative steps followed. On 13 March 2024, by 523 votes to 46, with 49 abstentions, the EP adopted the compromise text at first reading and therefore, under the ordinary legislative procedure, on 21 May 2024 the Council approved the EP position and the *Act* itself, which was signed on 13 June 2024 and entered into force on the 1st of August. It applies only to areas within EU law and provides exemptions for systems used exclusively for military, defence, or national security and for research purposes.

Conclusion

Despite the “reductions” due to the interinstitutional compromise, the EP’s influence on the Regulation is quite prominent. The use of a real-time remote biometric identification system for the purposes of law enforcement in publicly accessible spaces is now authorised only if the relevant authority has

³⁰ The consulted compromise text expresses the point of view of the Council’s Presidency.

completed a fundamental rights impact assessment. Besides, its use remains limited to what is strictly necessary both temporally and geographically. But most of all, «no decision producing an adverse legal effect on a person should be taken based solely on the output of the remote biometric identification system» (Regulation EU 2024/1689). The regulation also lays down clear obligations for high-risk AI systems, due to their significant potential harm to health, safety, fundamental rights, environment, democracy and the rule of law: they must assess and reduce risks, maintain use logs, be transparent and accurate, and ensure human oversight. For a high-risk AI system to be deployed by entities providing public services, its fundamental rights impact needs to be assessed. Besides, high-risk AI systems, as well as certain users of a high-risk AI system that are public entities, need to be registered in the relevant EU database, while users of an emotion recognition system must inform natural persons when they are being exposed to such a system. Finally, citizens can submit complaints about AI systems and receive explanations about decisions based on high-risk AI systems that affect their rights. GPAI systems, and the models such as ChatGPT they are based on, must meet transparency requirements, including compliance with EU copyright law. At the same time, artificial or manipulated images, audio, or video content (“deep fakes”) must be clearly labelled as such. It seems, therefore, proper to state that the current EU regulation on AI has absorbed, although not entirely embraced, the “political” and “ideal” approach adopted by the EP in the legislative process. Although in some cases, it is in a softened form, the core of democracy-focused amendments proposed by the European assembly has indeed been incorporated, especially as far as the scope of the Act is concerned. As for the reductions safeguarding national sovereignty, security, and interests, they can be ascribed to the intergovernmental legislative

and institutional structure of the EU, and therefore they couldn't be either avoided or opposed. If the risk-based approach – the higher the risk to cause harm to society, the stricter the rules – is the major contribution of the Commission to this EU legal milestone, it can be argued that the human-centric, fundamental rights-protecting, democracy-safeguarding purpose, which surely represents the *AI Act* most innovative feature, stems from the EP legislative action.

References

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M. & Floridi, L. (2018). Artificial intelligence and the “good society”: the US, EU, and UK approach. *Science and Engineering Ethics*, 24(2), 505–28.

Cinquanta, R. (2022). La Conferenza sul futuro dell'Europa: potenzialità e limiti di un esperimento di democrazia partecipativa sovranazionale. *lacittadinanzaeuropeaonline*, 2, 81-94.

Conference on the future of Europe (2022). Report on the final outcome. <https://www.europarl.europa.eu/resources/library/media/20220509RES29121/20220509RES29121.pdf>.

Council of the European Union (2022). Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [...] – General approach. <https://www.consilium.europa.eu/en/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights/>.

Council of the European Union (2024). Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [...] – Analysis of the final compromise text with a view to agreement. Available: <https://www.consilium.europa.eu/en/press/press-releases/2024/05/21/artificial-intelligence-ai-act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/>.

Council of the European Union, European Parliament (2024). Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). <https://www.consilium.europa.eu/en/press/press-releases/2024/05/21/artificial-intelligence-ai-act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/>.

Declaration of cooperation on AI (2018). <https://digital-strategy.ec.europa.eu/en/news/eu-member-states-sign-cooperate-artificial-intelligence>.

EC (2018a). Communication «Artificial Intelligence for Europe». <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN>.

EC (2018b). Communication «Coordinated Plan on Artificial Intelligence». <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018DC0795>.

EC (2019). Communication «Building Trust in Human-Centric Artificial Intelligence». <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52019DC0168&qid=1730968172043>.

EC (2020b). Communication «Shaping Europe’s digital future». <https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A52020DC0067>.

EC (2020c). Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0064>.

EC (2020d). White Paper on Artificial Intelligence – A European approach to excellence and trust. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0065&qid=1730968388349>.

EC (2021a). Communication «Fostering a European approach to Artificial Intelligence». <https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM%3A2021%3A205%3AFIN>.

EC (2021b). Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (AI Act) and amending certain Union legislative acts. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>.

EC (2022). Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive). <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0496>.

EP (2017). Resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics.

https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html.

EP (2020a). Decision of 18 June 2020 on setting up a special committee on artificial intelligence in a digital age, and defining its responsibilities, numerical strength and term of office. https://www.europarl.europa.eu/doceo/document/TA-9-2020-0162_EN.html.

EP (2020b). Resolution of 20 October 2020 on intellectual property rights for the development of artificial intelligence technologies. https://www.europarl.europa.eu/doceo/document/TA-9-2020-0277_EN.html.

EP (2020c). Resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence. https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_EN.html.

EP (2020d). Resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies. https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_EN.html.

EP (2021a). Resolution of 20 January 2021 on artificial intelligence: questions of interpretation and application of international law in so far as the EU is affected in the areas of civil and military uses and of state authority outside the scope of criminal justice. https://www.europarl.europa.eu/doceo/document/TA-9-2021-0009_EN.html.

EP (2021b). Resolution of 6 October 2021 on artificial intelligence in criminal law and its use by the police and judicial authorities in criminal matters. https://www.europarl.europa.eu/doceo/document/TA-9-2021-0405_EN.html.

EP (2022). Resolution of 3 May 2022 on artificial intelligence in a digital age. https://www.europarl.europa.eu/doceo/document/TA-9-2022-0140_EN.html.

EP (2024). Legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html.

EP Committee on the Internal Market and Consumer Protection, Committee on Civil Liberties, Justice and Home Affairs, (2023). Report on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act). https://www.europarl.europa.eu/doceo/document/A-9-2023-0188_EN.html.

EP Special Committee on Artificial Intelligence in the Digital Age, (2022a). Activity Report following AIDA Mandate 2020-2022.

EP Special Committee on Artificial Intelligence in the Digital Age (2022b). Report on artificial intelligence in a digital age. https://www.europarl.europa.eu/doceo/document/A-9-2022-0088_EN.html.

European Commission (henceforth EC) (2020a). Communication «A European Strategy for Data». <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52020DC0066>.

European Parliament (henceforth EP) (2023), Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html.

European Parliamentary Research Service (2023). Artificial intelligence, democracy and elections. [https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI\(2023\)751478](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2023)751478).

European Parliamentary Research Service (2024a). Artificial Intelligence Act. [https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI\(2021\)698792](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2021)698792).

European Parliamentary Research Service (2024b). The arrival of e-voting and campaign technologies in Europe: Promise, perils and preparedness. [https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI\(2024\)762321](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2024)762321).

High-Level Expert Group on Artificial Intelligence (2019a). Ethics Guidelines for Trustworthy Artificial Intelligence. Available: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

High-Level Expert Group on Artificial Intelligence, (2019b), Policy and investment recommendations for trustworthy Artificial

Intelligence. Available: <https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence>.

High-Level Expert Group on Artificial Intelligence (2020). Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment. <https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>.

Laux, J., Wachter, S., Mittelstadt, B. (2022). Trustworthy Artificial Intelligence and the European Union AI Act: on the conflation of trustworthiness and the acceptability of risk, <http://dx.doi.org/10.2139/ssrn.4230294>.

Moccia, L. (2021). «Cittadini Uniti d'Europa». In Zucca F., Cinquanta R. (a cura di), *La paura del Leviatano europeo*, *De Europa*, Special Issue 2021, Collane@unito, 375-398.

Mügge, D. (2024). EU AI sovereignty: for whom, to what end, and to whose benefit? *Journal of European Public Policy*, 31(8), 2200-2225.

Nardocci, C. (2024). Artificial intelligence at the crossroads between the European Union and the Council of Europe: who safeguards what how? *Italian Journal of Public Law*, 16(1), 165-196.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free

movement of such data [...] (General Data Protection Regulation). <https://eur-lex.europa.eu/eli/reg/2016/679/oj>.

Regulation (EU) 2023/988 of the European Parliament and of the Council of 10 May 2023 on general product safety. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2023.135.01.0001.01.ENG&toc=OJ%3AL%3A2023%3A135%3ATOC).

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence [...] (Artificial Intelligence Act). *Official Journal of the European Union*, L, 1689, 12 July 2024.

Roberts, H., Cowls, J., Casolari, F., Morley, J., Taddeo, M. & Floridi, L. (2021a). Safeguarding European values with digital sovereignty: an analysis of statements and policies, *Internet Policy Review*, 10(3). 1–26.

Roberts, H., Cowls, J., Hine, E., Mazzi, F., Tsamados, A., Taddeo, M. & Floridi, L. (2021b). Achieving a “good AI society”: comparing the aims and progress of the EU and the US, *Science and Engineering Ethics*, 27(6), 1–25.

Roberts, H., Cowls, J., Hine, E., Morley, J., Wang, V., Taddeo, M. & Floridi, L. (2023), Governing artificial intelligence in China and the European Union: comparing aims and promoting ethical outcomes, *The Information Society*, 39(2), 79-97.

Ulnicane, I. (2021). Artificial Intelligence in the European Union: policy, ethics and regulation. In Hoerber, T., Cabras, I., Weber, G., (Eds.), *Routledge Handbook of European Integrations*. Routledge. 254-269.

Van Noordt, C., Misuraca, G. (2022), Artificial intelligence for the public sector: results of landscaping the use of AI in government across the European Union, *Government Information Quarterly*, 39(3). <https://doi.org/10.1016/j.giq.2022.101714>.

7. L'IA NELLA GESTIONE DELLE FRONTIERE DELL'UNIONE EUROPEA: UN APPROCCIO ANTROPOCENTRICO PER TUTTI?

di Giulia Maria Gallotta*

Abstract: Il saggio analizza la definizione dell'approccio dell'UE allo sviluppo dell'intelligenza artificiale, dalla risoluzione del Parlamento europeo del 2017 al recente AI act e del concetto di smart borders nella politica migratoria comune. Obiettivo è mettere in luce come strumenti analoghi si muovano in direzioni nei fatti opposte, creando livelli di tutela dei diritti profondamente diversi per i cittadini e i migranti legali, da un lato, e per i migranti cosiddetti illegali dall'altro.

Parole chiave: Unione europea, smart borders, intelligenza artificiale, migrazione, dati biometrici

Abstract: The essay analyses the forming lines of the EU's approach to the development of artificial intelligence, from the 2017 European Parliament resolution to the recent AI Act and the concept of smart borders in the common migration policy. The aim is to highlight how similar instruments move in fact in opposite directions, creating profoundly different levels of rights protection for legal citizens and migrants on the one hand, and for so-called illegal migrants on the other.

* Dipartimento di Scienze della Formazione, Psicologia e Comunicazione, Università degli Studi di Bari "Aldo Moro", ORCID 0000-0003-3511-2312, giuliamaria.gallotta@uniba.it.

Accettato Dicembre 2024 - Pubblicato Aprile 2025.

Keywords: European Union, smart borders, artificial intelligence, migrants, biometrics

Introduzione

Seguire gli studenti nel loro lavoro di tesi è talvolta l'occasione per apprendere cose nuove o per rileggere in modo diverso argomenti su cui si lavora già da un po'. Questo accade specialmente quando il loro entusiasmo feconda ricerche che sembrano arrivate ad un punto morto, sia perché gli sviluppi più interessanti richiedono competenze estremamente specialistiche in ambiti nettamente al di fuori della propria disciplina sia perché si indossano facilmente le vesti dei profeti inascoltati. È quello che mi è capitato riguardo ai temi di questo saggio. Sempre più spesso, gli sviluppi in materia di intelligenza artificiale (IA) applicati al governo dei migranti sono troppo specialistici, poiché avvengono in ambiti di ingegneria informatica avanzata rispetto ai quali non ho alcuna competenza. D'altro canto, gli studi sugli effetti dell'uso dell'IA nella gestione delle politiche migratorie nazionali o europee sono tristemente in linea con quanto gli studiosi ipotizzano già da qualche anno. Per quanto riguarda la gestione dei migranti da parte dell'Unione europea, questa ha da poco approvato i regolamenti che danno attuazione al *Nuovo Patto sulla migrazione e l'asilo*, proposto dalla presidente della Commissione europea U. von der Leyen nel settembre 2020. Il quadro normativo europeo appare completo ed entrerà pienamente in vigore nel 2026, al di là di dichiarazioni di alcuni governi nazionali sul rimetterlo in discussione. Queste appaiono più ad uso di posizionamento politico interno che foriere di un'effettiva riapertura di negoziati che, a livello europeo, sono già stati estremamente lunghi e

difficili¹. Lo stesso programma strategico, presentato dalla von der Leyen al Parlamento europeo (PE) nel luglio 2024, in occasione della sua candidatura per un nuovo mandato come Presidente della Commissione, dedica alla politica della migrazione due brevi paragrafi verso la metà del documento e solo per ribadire come una solida politica di controllo delle frontiere comunitarie, e quindi la lotta alla migrazione irregolare e la cooperazione con i Paesi di origine e di transito dei migranti per agevolarne i rimpatri, resti uno dei pilastri a garanzia della libera circolazione dei cittadini comunitari all'interno dello spazio europeo senza frontiere (von der Leyen, 2024). Anche se, a questo proposito, è più interessante quello su cui si tace². Poi, uno studente chiede di sviluppare una tesi che parte dall'analisi di un videogioco "Papers, please!" nel quale il giocatore riveste il ruolo di un'anonima guardia di frontiera di un immaginario Paese autocratico. Della guardia si sa solo che ha una famiglia numerosa e che il suo stipendio aumenta o diminuisce in funzione dei respingimenti che effettua. L'aspetto interessante è che il giocatore conosce sia il grado di regolarità dei documenti che gli vengono presentati sia le storie di coloro che glieli sottopongono. Questo significa che questi è chiamato a

¹ Mi riferisco al ministro dell'Interno francese, Bruno Retailleau che ha dichiarato di voler rinegoziare i regolamenti attuativi del *Nuovo Patto per la migrazione e l'asilo*. In linea con la maggior parte dei commentatori, mi limito a sottolineare che il ministro appartiene al partito repubblicano, fortemente in crisi a causa della netta affermazione del partito di M. Lepen, noto per le sue dure posizioni contro i migranti. E a ricordare che i negoziati sui regolamenti attuativi del *Nuovo Patto sulla Migrazione e l'asilo* hanno richiesto oltre tre anni per arrivare a conclusione. Di diverso tenore è la decisione del primo ministro polacco Donald Tusk di rifiutare di applicare il *Nuovo Patto* per introdurre criteri fortemente selettivi di accesso al proprio Paese. Questa si basa sul fatto che la Polonia accoglie larghissima parte di rifugiati ucraini e che i ricorrenti afflussi di migranti dalla Bielorussia sarebbero parte di una guerra ibrida, orchestrata dalla Russia. Cfr. «Parigi si dice disposta alla revisione del Patto su migrazione e asilo e alla riforma di Schengen», *Euractiv*, 25.09.2024.; «Il governo polacco di Tusk adotta ufficialmente una nuova strategia sulle migrazioni», *Euractiv*, 16.10.2024.

² Tornerò su questo punto *infra*.

comporre in un delicato equilibrio la propria (eventuale) volontà di aiutare soggetti che sa vulnerabili a dispetto dei documenti non proprio in regola con le decurtazioni che questo comporta sul suo magro stipendio, talvolta vedendosi costretto ad accettare criminali e a respingere persone non in regola per le quali l'ingresso rappresenta l'unica possibilità di sopravvivenza, pur di non essere escluso dal *game*. Tornerò su questo videogioco nelle conclusioni. Prima mi pare opportuno partire dallo sviluppo di due politiche comunitarie: la definizione di un approccio comune allo sviluppo dell'intelligenza artificiale (IA) e il concetto di *smart borders* come asse della politica migratoria dell'Unione europea. Entrambe sono nate nello stesso lasso di tempo, fra il 2016 e il 2017, e si sono sviluppate in modo indipendente l'una dall'altra. Gli sviluppi recenti invitano, però, a valutarne attentamente le reciproche implicazioni.

1. L'approccio antropocentrico all'IA dell'Unione europea

A portare la questione dello sviluppo dell'IA sotto i riflettori comunitari è una risoluzione del PE del 2017 sulla robotica, agli inizi degli scandali sulle attività di profilazione condotte attraverso sistemi di IA nelle campagne elettorali per il referendum sulla Brexit e per l'elezione di D. Trump alla presidenza degli Stati Uniti. Negli stessi anni cominciano anche i primi utilizzi di robot infermieri in alcuni ospedali. Di fronte a questi sviluppi, il Parlamento comincia a porre la questione delle conseguenze che lo sviluppo di macchine capaci di agire al posto degli uomini e di svolgerne le mansioni può avere. Tanto più quando queste vengono utilizzate in ambiti delicati, nei quali la dimensione dell'interazione umana gioca un ruolo preminente, come le attività di cura, la salute o l'adozione di decisioni che possono avere conseguenze

importanti sui loro destinatari. L'auspicio che il PE formula è che gli sviluppi in questo ambito siano basati sui principi della «sicurezza delle persone e della loro salute, della libertà, la vita privata, l'integrità, la dignità, dell'autodeterminazione e la non discriminazione nonché della protezione dei dati personali» (Risoluzione PE, 2017: 10). Nel giro di un anno anche la Commissione europea e il Consiglio dei ministri cominciano ad articolare quello che è stato definito un approccio antropocentrico all'IA. Più precisamente, è la Commissione europea a definirne progressivamente i contenuti attraverso tre comunicazioni, adottate in rapida successione fra il 2018 e il 2020. La prima si concentra sulle conseguenze in termini di crescita economica che lo sviluppo dell'IA può generare e sui ritardi degli investimenti europei nel settore rispetto ai principali competitors internazionali, USA e Cina in particolare. In questo senso, la Commissione sottolinea la necessità che l'azione dell'Unione e dei suoi Stati membri si sviluppi lungo due linee parallele. Da un lato, si tratta di promuovere misure di formazione dei lavoratori e percorsi scolastici che consentano ai cittadini europei di acquisire le conoscenze e le abilità necessarie per inserirsi in modo proficuo nella nuova economia dei dati. L'ambizione è anche quella di orientarne gli sviluppi, grazie all'elevata qualità del capitale umano europeo. D'altro canto, la Commissione sottolinea come sia essenziale stabilire un quadro europeo di regolazione dell'IA che, da un lato, garantisca che il suo sviluppo sarà rispettoso ed in linea con i valori fondamentali dell'UE e, dall'altro, renda quest'ultima capace di orientarne gli sviluppi a livello mondiale, quasi a rivendicare per l'Unione una sorta di ruolo di potenziale *standard setter* (European Commission, 2018). Quali siano i valori fondamentali che la Commissione pone alla base dello sviluppo europeo dell'IA viene esplicitato in una breve comunicazione del 2019, dove l'istituzione richiama la lettera dell'art. 2 del trattato di

Lisbona, secondo il quale l'UE si fonda sui «valori del rispetto della dignità umana, della libertà, della democrazia, dell'uguaglianza, dello Stato di diritto e del rispetto dei diritti umani, compresi i diritti delle persone appartenenti a minoranze» (European Commission, 2019:2). Il rispetto di questi valori è per la Commissione un prerequisito, al quale si aggiungono precise caratteristiche di funzionamento che i sistemi di IA dovranno rispettare, in modo da consentire ai cittadini europei di fidarsene e di appropriarsene a livello sia personale che economico. Fra questi, riveste un ruolo centrale la possibilità di intervento umano nei processi decisionali, la riservatezza dei dati e la protezione dei principi di «diversity, non-discrimination and fairness» (ibid.: 3) e la spiegabilità dei sistemi di IA, ossia la capacità di giustificare i risultati generati dai rispettivi algoritmi. Cosa questo significhi è ben articolato nella comunicazione del 2020. Qui, il punto di partenza è la natura potenzialmente opaca dell'IA, che i suoi rapidi sviluppi tecnologici mettono in evidenza. L'opacità riguarda, da un lato, le possibilità di comprendere il funzionamento e la logica operativa degli algoritmi che regolano il processo di apprendimento dei sistemi di IA. Dall'altro, esiste la possibilità che le enormi quantità di dati che questi processano, insieme alla capacità dei sistemi stessi di individuare associazioni fra i dati o scorciatoie per raggiungere lo scopo loro assegnato al di fuori di quanto previsto dai programmatori, generi processi decisionali e/o decisioni non trasparenti o che riflettono discriminazioni di vario tipo. Lo stesso vale per attività criminali che possono essere condotte grazie all'uso di queste stesse tecnologie. In questo scenario, la Commissione sottolinea come non si tratti solo di promuovere attività di ricerca di alto profilo, la diffusione dell'IA presso le piccole e medie imprese (pmi) e partnership pubblico-privato per finanziare la ricerca sull'IA e la sua diffusione, quello che viene definito come un «ecosystem of excellence» (European Commission,

2020:3). L'obiettivo è creare un sistema normativo che non si limiti a proteggere «human dignity and privacy» (European Commission 2020:2) dei cittadini europei ma che fissi standard di progetto e operativi per qualunque progettista/fornitore/distributore/utilizzatore di sistemi di IA che voglia farne uso o venderli nell'UE. In questo senso, la Commissione sottolinea come i dati che alimentano ed allenano i sistemi di IA debbano rispettare la privacy dei soggetti dai quali provengono, come debba esserne verificata la neutralità rispetto a pregiudizi di ogni genere e debba essere sempre prevista la possibilità di una verifica umana, che si tratti della correzione dei dati stessi e/o dei risultati ottenuti, specialmente nei casi di decisioni algoritmiche che possono avere conseguenze importanti per i singoli, come le decisioni in ambito sanitario, processuale o per l'erogazione di prestazioni sociali o prestiti bancari. Soprattutto, tutti i sistemi di IA devono essere comprensibili e i processi che generano i loro risultati finali devono essere spiegabili (ibid.). Nell'arco di due anni, con queste tre comunicazioni, la Commissione traccia dunque le linee ed i principi che devono orientare lo sviluppo e la diffusione dell'IA nell'UE e che dovrebbero diventarne la bandiera nelle relazioni commerciali con i partner internazionali. Queste affondano le proprie radici nella necessità di conciliare la necessaria attenzione a tenere il passo con i ritmi di sviluppo di tecnologie che si evolvono molto più rapidamente di quanto i legislatori non riescano a regolarle e l'attenzione alle importanti conseguenze che queste stesse tecnologie avranno sulle persone, non solo in termini di obsolescenza delle loro abilità sul mercato del lavoro ma anche di potenziale sorveglianza delle attività più disparate e di influenza su opinioni personali e comportamenti socialmente rilevanti. La Commissione mira anche a preservare nello sviluppo tumultuoso delle nuove tecnologie informatiche il principio in base al quale queste, come tutta la tecnologia, debbano essere al servizio degli

individui e non lederne i diritti fondamentali di dignità e rispetto della privacy. Tutti i successivi atti normativi che la Commissione ha proposto in materia possono essere letti come applicazione di questi principi a casi specifici. E in questo senso, l'*AI act*, proposto con una comunicazione dell'aprile 2021, indica in modo chiaro quali sistemi di IA siano vietati nell'Unione e quali siano considerati ad alto rischio e di conseguenza soggetti al soddisfacimento di precisi requisiti in materia di progettazione e di controlli prima e dopo la loro immissione sul mercato interno europeo. Nella proposta della Commissione, sono vietati tutti quei sistemi di IA che, attraverso tecniche subliminali o sfruttando situazioni di fragilità dei soggetti ai quali si rivolgono, sono in grado di «distorcere materialmente il loro comportamento in un modo che causa o può causare un danno fisico o psicologico a quella o ad altre persone» (Commissione europea, 2021:art.5.a) o che possono essere usati a fini di valutazione sociale del comportamento di gruppi o individui (c.d. social scoring), l'uso di strumenti di identificazione biometrica remota in tempo reale in spazi ristretti, con specifiche eccezioni relative alla protezione della sicurezza o di vittime di reati. La definizione dei sistemi di IA ad alto rischio è, invece, vaga, poiché si basa sul loro grado di pericolosità, valutata a partire dalla loro capacità di influenzare negativamente il godimento dei diritti umani da parte dei singoli individui. La Commissione allega, però, alla proposta due liste che li elencano in maniera non esaustiva (Commissione europea, 2021, Allegato III). Queste includono, per esempio, i sistemi di IA che possono essere usati per stabilire la solvibilità dei singoli o la loro rispondenza ai criteri per ricevere prestazioni sociali o la loro valutazione nell'ambito di istituti di istruzione/formazione o in processi criminali o, in relazione alla gestione dei migranti, la valutazione della loro rispondenza ai criteri per ottenere la protezione internazionale o la veridicità delle dichiarazioni e dei

documenti prodotti a sostegno delle domande stesse (ib.) Per questi sistemi, sono previste precise procedure che ne regolano la fornitura e l'uso sul mercato interno europeo e modalità per il monitoraggio del loro funzionamento. Un articolo è anche dedicato ai sistemi di IA che interagiscono con le persone. Questo stabilisce l'obbligo che queste ultime siano messe a conoscenza della natura artificiale dell'interlocutore, specie quando quest'ultimo può registrare e identificare emozioni o generare contenuti *deep-fake* (ibid.: art. 52). La proposta della Commissione punta a delimitare il perimetro all'interno del quale l'IA si svilupperà nell'Unione europea. Questa potrà espandere il proprio potenziale all'interno di precisi criteri di rispetto per i diritti degli individui, che sono la carta di identità dell'Unione e di trasparenza per quanto riguarda l'uso e la trasparenza dei suoi algoritmi operativi³.

Fra la proposta della Commissione e l'adozione dell'*AI act* sono passati circa tre anni, due dei quali sono stati marchiati dalla pandemia di Covid e dal suo strascico di proliferazione di attività di ogni genere svolte online. Questo ha significato la moltiplicazione esponenziale dei dati in circolazione in rete, la loro disponibilità per i fini più disparati ma anche una maggiore e più diffusa familiarità dei cittadini di ogni età con l'uso dei supporti informatici (smartphone e via dicendo) in ogni genere di attività. In questo senso, il testo definitivo del regolamento sull'IA reca i segni evidenti delle modifiche, negoziate dal PE nella procedura di adozione congiunta con il Consiglio dei Ministri, e che vanno nella direzione di un irrigidimento del quadro normativo complessivo, nel tentativo di ridurre i margini di discrezionalità dei governi nazionali e la conseguente tentazione del *free-riding* (Novelli *et al.*, 2023).

³ A questo proposito, è il caso di sottolineare come l'approccio antropocentrico all'IA dell'UE sia stato oggetto di una certa ironia da parte degli studiosi, specialmente del mondo accademico statunitense, che lo giustappongono alle strategie di sviluppo molto più aggressive di Cina e Stati Uniti, ad oggi gli unici veri contendenti nella lotta per la supremazia digitale (Rida Nour, 2019).

Innanzitutto, è appena il caso di sottolineare che l'atto comunitario che definisce il quadro normativo per l'uso e lo sviluppo dell'IA è un regolamento, ossia un atto immediatamente applicabile, che può essere utilizzato dai cittadini europei nelle loro controversie davanti a giudici nazionali di qualunque ordine e grado⁴, e che sarà applicabile non solo agli sviluppatori/distributori/utilizzatori/venditori europei ma a quelli di qualunque nazionalità, purché i loro prodotti basati o contenenti elementi di IA siano utilizzati all'interno dell'Unione o abbiano semplicemente conseguenze al suo interno (Regolamento UE, 2024). Il testo è troppo ampio perché possa tentarne anche solo una semplice disamina. Mi concentrerò sugli aspetti più interessanti rispetto ai miei ambiti di ricerca. L'attenzione allo sviluppo di un'IA antropocentrica rimane ma, rispetto alla proposta della Commissione, al di là dei richiami alla tutela dei diritti fondamentali e del diritto alla protezione dei dati dei cittadini europei, la sua declinazione più articolata è formulata in materia di diritti del lavoro, rispetto ai quali l'adozione dell'IA deve essere neutra ossia non peggiorativa dei diritti dei lavoratori come riconosciuti a livello europeo. È questo un tentativo di regolare le condizioni di sostanziale sfruttamento in cui versa oggi larga parte del lavoro salariato, specie quello impiegato nella gig economy, o nelle grandi piattaforme che vendono servizi/prodotti online così come di evitare tentativi di controllo/profilazione sul luogo di lavoro. Si tratta, però, di un tentativo non vincolante in quanto è inserito nel preambolo del regolamento (Regolamento UE, 2024: par.9). Importante (forse illusorio?) è anche l'impegno a promuovere uno sviluppo dell'IA che sia rispettoso dell'ambiente, soprattutto quando è ormai noto che l'impatto di queste tecnologie sulla natura è decisamente importante in termini di inquinamento e di consumo energetico (ivi, par. 27). L'irrigidimento del

⁴ Questo in virtù di una consolidata giurisprudenza della Corte di Giustizia dell'UE, che risale alla sentenza nota come *Politi* del dicembre 1971.

regolamento rispetto alla proposta della Commissione balza agli occhi nei primi articoli, dove le definizioni delle attività e delle figure che progettano/producono/vendono/usano i sistemi di IA è estremamente precisa. Lo stesso può dirsi per il complesso sistema di controlli e certificazioni sui quali la progettazione, lo sviluppo e la messa in uso dei sistemi di IA deve basarsi. Questo dovrebbe contribuire a ridurre il rischio di malfunzionamento dei sistemi o di utilizzo di set di dati per il loro allenamento che risulti viziato da pregiudizi di genere, di razza e così via. Su questo punto, è stato fatto osservare come si tratti di requisiti che, da un lato, rischiano di essere eccessivamente onerosi per le imprese di piccole e medie dimensioni, con la conseguenza di ritardarne o bloccarne l'adozione di tecnologie altamente innovative. D'altro canto, non sembra essere un sistema pur rigoroso di certificazioni a poter contribuire a sanare/compensare eventuali danni causati dai sistemi di IA ai diritti dei singoli utilizzatori (Finocchiaro, 2023). Dal mio punto di vista, le novità più interessanti sono due e riguardano entrambe i sistemi di IA ad alto rischio. Innanzitutto, la definizione che ne viene data nell'art.6 del regolamento è decisamente poco comprensibile per un profano di diritto comunitario della libera circolazione dei beni: questa è, infatti, articolata solo in riferimento alle normative europee in materia di armonizzazione e di conformità di questi ultimi per quanto riguarda la loro immissione in commercio (Regolamento UE, 2024: art. 6). Altre indicazioni sono ricavabili *a contrario* dalla successiva indicazione di cosa rende un sistema di IA non ad alto rischio, ossia il fatto che non presenti un «rischio significativo di danno per la salute, la sicurezza o i diritti fondamentali delle persone fisiche, anche nel senso di non influenzare materialmente il risultato del processo decisionale» (ivi: art. 6.3). Questo si verifica quando i compiti dell'algoritmo sono limitati e non sono finalizzati a sostituire in toto la valutazione umana in un processo

valutativo/decisionale. Quali siano concretamente i sistemi di IA ad alto rischio viene dettagliato nell'allegato III al regolamento⁵. Alcuni studiosi hanno sottolineato come nell'indicarli questo tenga insieme nei fatti due criteri distinti, ossia quello della capacità dei sistemi di IA di creare problemi se usati in modo non etico e quello della loro capacità di causare danni gravi se il loro funzionamento viene sospeso (Floridi, 2021). Altri mettono in rilievo come la definizione di un sistema ad alto rischio dipenda alla fine da una decisione della Commissione, affiancata dagli Stati membri, e non da valutazioni dei soggetti che lo progettano/distribuiscono/utilizzano. Questo, insieme alla macchinosità del meccanismo per il loro aggiornamento, condurrebbe alla creazione di un sistema troppo rigido e modellato sul presente, poco capace di reagire al dinamismo proprio dei ritmi di sviluppo dell'IA (Finocchiaro, 2023). La seconda novità riguarda le procedure, previste dietro esplicita richiesta del PE, per tutelare i diritti umani da violazioni che possono nascere dall'uso di sistemi di IA ad alto rischio. Al di là della retorica sulla loro tutela, nei fatti il meccanismo di tutela previsto sembra ridursi a controlli preventivi del grado di pericolosità dei sistemi stessi, effettuati dai loro utilizzatori, ossia coloro che utilizzano sistemi di IA ad alto rischio sotto la propria autorità (Regolamento UE, 2024: art. 3.4). Il rischio di conflitti di interesse nella valutazione è stato messo in rilievo dagli studiosi, come la sostanziale assenza di meccanismi per rimediare a casi in cui le violazioni si manifestino. L'unica norma in merito prevede il diritto del soggetto leso a ricevere una spiegazione sul modo in cui il sistema è arrivato al risultato. Quanto alla previsione di un controllo umano del processo decisionale del sistema ad alto rischio nel suo concreto svolgersi, questo appare estremamente

⁵ Per completezza dell'argomentazione, mi limito a sottolineare come l'art. 5 dell'*AI act*, dedicato ai sistemi di IA vietati, riprenda in toto e senza sostanziali variazioni la formulazione della Commissione.

difficile da realizzare (Contissa e Galli, 2024; Cirone, 2024). A questa vaghezza sui diritti e sulla loro attivabilità da parte dei cittadini, si aggiunge quello che personalmente trovo l'elemento più interessante dal punto di vista della politica migratoria dell'UE. Infatti, sono considerati come ad alto rischio anche i sistemi di identificazione biometrica, ossia quei sistemi che si basano sul confronto dei dati biometrici e/o delle immagini facciali dei singoli per verificarne l'identità, i sistemi utilizzati a fini di categorizzazione biometrica, ossia finalizzati ad inserire i soggetti in banche dati «che classificano individualmente le persone fisiche sulla base dei loro dati biometrici per trarre deduzioni o inferenze in merito a razza, opinioni politiche, appartenenza sindacale, convinzioni religiose o filosofiche, vita sessuale o orientamento sessuale» (ivi:art.5.g), i sistemi per il riconoscimento delle emozioni. Inoltre, per quanto riguarda i migranti regolari e irregolari, sono considerati ad alto rischio i sistemi di IA che verificano l'identità delle persone, la veridicità delle loro affermazioni o che valutano il rischio che il soggetto pone dal punto di vista della sicurezza o della regolarità del suo soggiorno. La natura ad alto rischio di un sistema di IA ne subordina l'utilizzo a procedure di autorizzazione da parte delle autorità giudiziarie nazionali che restano, però, nella discrezionalità degli Stati membri, essendo previste nel preambolo (ivi: al. 22). Quello che più mi colpisce, a questo proposito, è lo strabismo fra la disciplina dell'IA e la politica migratoria dell'UE.

2. L'IA nella politica migratoria dell'UE

Tutto comincia nel 2011, quando la Commissione europea propone la creazione di un'agenzia tecnica alla quale affidare la gestione dei database necessari alla realizzazione dello spazio Schengen di libera circolazione delle persone. Inizialmente, si

tratta del Sistema Informativo Schengen (SIS) che collega le banche dati degli Stati membri relative alle persone condannate o oggetto di indagini nello spazio senza frontiere, del Sistema di Informazione Visti (VIS) che raccoglie i dati di coloro che fanno richiesta di un visto per entrare nel territorio comunitario, delle richieste stesse e del loro esito, e di EURODAC, il sistema che raccoglie le foto e le impronte digitali dei migranti irregolari che arrivano alle frontiere dell'Unione. Nasce così eu-LISA, l'agenzia dell'Unione europea per la gestione dei sistemi IT su larga scala nello spazio di libertà, sicurezza e giustizia. Suo compito è garantire il corretto funzionamento dei database comunitari 24h su 24, 7 giorni su 7, e la progettazione, sviluppo e gestione di tutti i sistemi e le reti informatiche che le istituzioni comunitarie, gli sviluppi tecnologici o le possibili sinergie fra queste richiedano o rendano necessari (Regolamento 2011: art. 1; Glouftios, 2021). Il problema è che, come recita un noto adagio popolare, l'occasione fa l'uomo ladro. E il possesso di importanti quantità di dati genera bulimia. Trascorso il tempo necessario per mettere materialmente in piedi eu-LISA e rodarne i sistemi di funzionamento, nel 2016 la Commissione adotta una comunicazione con la quale lancia la strategia per trasformare i confini comunitari in *smart borders*. È appena il caso di ricordare che già da un anno l'UE è alle prese con la c.d. crisi dei migranti, ossia l'arrivo sulle coste greche, italiane e spagnole di numeri significativi di migranti in fuga dall'instabilità politica ed economica dei Paesi africani e della riva nord del Mediterraneo. La gestione di questa emergenza ha evidenziato sia la sostanziale incapacità dei Paesi europei di primo approdo di accogliere in modo dignitoso i migranti⁶ e di processarne in tempi

⁶ È appena il caso di ricordare che in una sentenza del 2011 la Corte di Giustizia dell'UE ha definito le condizioni degli hotspot sulle isole greche come assimilabili al "trattamento inumano e degradante", vietato dalle Convenzioni ONU e dalla Carta dei diritti fondamentali dell'UE. Cfr. CGE, N. S., C-411/10 e C-493/10, punto 86, 21 dicembre 2011.

rapidi le domande di protezione internazionale sia la fondamentale mancanza di solidarietà degli Stati membri verso quelli di primo approdo. In questo scenario, la Commissione traslascia la gestione dell'emergenza per concentrarsi sulle criticità e sulle lacune delle banche dati comunitarie esistenti, relative all'attraversamento delle frontiere. Queste scontano la scarsa qualità dei dati talvolta inseriti dalle autorità nazionali preposte, la loro frammentazione e la mancanza di strumenti dedicati ai problemi che l'accesso legale di cittadini terzi al territorio comunitario può comportare (Commissione europea, 2016). In questo senso, la strategia proposta dalla Commissione si articola su due livelli: da un lato, si tratta di completare il sistema di dati sui viaggiatori regolari sia aggiornando quelli raccolti nelle banche dati esistenti sia creandone di nuove. In questo senso, la Commissione propone di ampliare la banca dati del SIS ai dati biometrici delle persone oggetto di ricerche da parte delle autorità giudiziarie dell'UE, di coloro che sono destinatari di un divieto di ingresso o che non hanno ottemperato ad un ordine di lasciare il territorio comunitario. Per quanto riguarda la creazione di nuovi database, la Commissione propone di registrare ed archiviare le informazioni sui viaggiatori extra-comunitari che varcano le frontiere dell'UE in modo legale. In questo senso, l'Entry/Exit System (EES) sarà dedicato alla raccolta e all'archiviazione dei dati anagrafici, biometrici e di viaggio di tutti coloro che entrano ed escono dall'Unione, a prescindere dalla necessità di un visto, per monitorarne la persistente legalità del soggiorno o casi di permanenza non autorizzata (c.d. overstaying). Lo European Travel Information and Authorization System (ETIAS) raccoglierà, invece, le richieste di autorizzazione preventiva all'ingresso nell'UE e le informazioni sul viaggio stesso che i cittadini di Stati terzi ai quali non è richiesto un visto dovranno ottenere e gli esiti delle relative domande (ivi). D'altro canto, la Commissione si preoccupa di

facilitare l'accesso a questa massa importante di informazioni da parte di coloro che ne hanno materialmente bisogno per svolgere le proprie attività di sorveglianza e/o di contrasto. In questo senso, viene proposta la creazione di un sistema di interoperabilità, ossia un'architettura di collegamento fra tutte le banche dati comunitarie relative a soggetti che attraversano in maniera più o meno regolare le frontiere comunitarie. Questa prevede lo sviluppo di un'interfaccia di ricerca unica e di un sistema di verifica incrociata delle corrispondenze fra i dati biometrici archiviati (ivi). E il regolamento del 2019, che dà attuazione a questa proposta, prevede la creazione di un portale di ricerca unico, che dia accesso a tutti i database comunitari in materia di sicurezza e migrazione, incluse le banche dati ECRIS-TCN⁷, di un servizio comune di confronto dei dati biometrici in queste archiviati, di un archivio comune dei dati relativi all'identità delle persone e di un sistema per la rilevazione di identità multiple (Regolamento 2019/817). Tutte queste proposte hanno fatto la loro strada. L'EES è operativo da novembre 2024, ETIAS e i meccanismi per l'interoperabilità delle banche dati comunitarie richiedono uno sforzo sia di progettazione da parte dei tecnici di eu-LISA sia di implementazione da parte degli Stati membri. Il regolamento EURODAC è stato modificato per includere i minori fra i soggetti ai quali possono essere prelevate le impronte digitali e per inserire le immagini digitali dei volti fra i dati biometrici da registrare per tutti i migranti irregolari che arrivino ai confini dell'UE (Regolamento UE 2024/1358). Di tutto questo non c'è traccia nel programma che la von der Leyen ha presentato al PE per la propria candidatura ad un nuovo mandato. *Smart borders* ed interoperabilità proseguono il

⁷ ECRIS-TCN è l'acronimo per European criminal records information system-Third Country Nationals e designa la rete di collegamento fra gli uffici giudiziari degli Stati membri per lo scambio di informazioni su condanne penali riportate da cittadini non comunitari in qualsiasi Paese dell'Unione.

loro cammino quasi sottotraccia, nel silenzio delle istituzioni e nel parallelo rumore di tutti quei gruppi/partiti di destra e destra estrema che mobilitano il consenso popolare nel nome della difesa dall'invasione di migranti e dalla minaccia per il benessere e le identità nazionali che la loro presenza rappresenta.

3. Alcune considerazioni critiche sugli smart borders

In effetti, la trasformazione dei confini europei in *smart borders* e i piani per l'interoperabilità dei database comunitari sollevano più di un problema. L'obiettivo più generale e decisamente condivisibile delle misure comunitarie era e resta quello di migliorare il livello di tutela della sicurezza dei suoi cittadini, senza che questo comporti una sorta di baratto fra sicurezza e tutela dei diritti individuali. Lo stesso regolamento sull'interoperabilità afferma solennemente che il

«trattamento di dati personali ai fini del presente regolamento non dà luogo a discriminazioni nei confronti delle persone fondate sul genere, sulla razza, sul colore della pelle o sull'origine etnica o sociale, sulle caratteristiche genetiche, sulla lingua, sulla religione o sulle convinzioni personali, sulle opinioni politiche o di qualsiasi altra natura, sull'appartenenza a una minoranza nazionale, sul patrimonio, sulla nascita, sulla disabilità, sull'età o sull'orientamento sessuale. Esso rispetta pienamente la dignità e l'integrità umana nonché i diritti fondamentali, compreso il diritto al rispetto della vita privata e alla protezione dei dati personali» (Regolamento 2019/817: art. 5).

Tutto a posto, allora? Non proprio. Meglio, non per tutti.

Una prima questione riguarda proprio la protezione dei dati personali dei migranti. Come sottolinea Martina Tazzioli in un

recente studio sui software utilizzati nelle isole greche per monitorare i movimenti dei migranti lì trattenuti attraverso l'uso di carte elettroniche sulle quali vengono accreditate le risorse minime per il loro sostentamento, la questione non è solo quali dati vengono registrati ed archiviati ma anche di «who has access to what and what actors can do through that» (Tazzioli, 2022:9). In effetti, i tecnici di eu-LISA progettano sistemi informatici e di IA per la gestione delle banche dati comunitarie e la loro interoperabilità ma nei fatti l'agenzia acquista spazi online e servizi da aziende informatiche private. Queste sono vincolate da norme rigorose sulla protezione dei dati ma c'è da chiedersi quanto queste vengano effettivamente rispettate. Il problema non è quello (ad oggi banale) della vendita dei dati anonimizzati dei migranti ad altri soggetti per fini commerciali ma l'uso che le stesse aziende appaltatrici possono farne. Gli studiosi stanno da tempo sottolineando come il trattamento dei dati dei migranti e sulle loro abitudini di consumo si trasformi sempre più spesso in materiale dal quale le aziende informatiche ricavano indicazioni per servizi sempre nuovi e sempre più mirati da proporre a clienti vecchi e nuovi (Lemberg-Pedersen 2024; Lemberg-Pedersen and Haioty, 2020). Questo si intreccia ad una seconda questione, quella di chi sia titolare della possibilità di avere accesso ai dati conservati nelle banche dati comunitarie. A questo proposito, già nel 2019 il rapporto di PICUM e Statewatch sulle conseguenze dell'interoperabilità dei database comunitari per i migranti irregolari sottolineava come la definizione dei soggetti autorizzati all'accesso nel regolamento del 2019, fosse troppo ampia e non garantisse dalla possibilità di accessi abusivi o da un uso degli esiti delle ricerche non in linea con le norme comunitarie o nazionali (PICUM 2019). Questo dando per scontato che in fase di rilevamento e registrazione dei dati le autorità nazionali siano rispettose della privacy dei soggetti dai quali li rilevano. Un'obiezione più generale

è quella ormai ricorrente sul progressivo inquadramento delle questioni della migrazione e dell'asilo in termini di sicurezza direi quasi predittiva. Matthias Leese ha sottolineato come lo sviluppo del potere sovrano degli Stati si sia storicamente sviluppato parallelamente al costante affinamento delle capacità delle autorità pubbliche di acquisire informazioni sull'identità e sui mezzi economici delle persone che risiedevano all'interno del loro territorio per ricavarne risorse attraverso i sistemi di imposizione fiscale e strumenti per la difesa attraverso la coscrizione obbligatoria (Leese, 2020). A partire dagli attentati dell'11 settembre 2001 e ancora di più ai giorni nostri, dopo che attentati di matrice islamica hanno insanguinato diverse città europee e nel clima di insicurezza generato da quei migranti che girano per l'Europa vivendo di espedienti nell'attesa di provare a raggiungere i Paesi che sono la vera destinazione del loro viaggio, a questo si è aggiunta la preoccupazione di individuare in anticipo sia chi può essere un potenziale criminale sia chi, pur entrato legalmente, potrebbe andare ad ingrossare i gruppi degli stranieri irregolari, prolungando il proprio soggiorno al di là del tempo consentito.

Come hanno sottolineato Philippa Metcalfe e Lina Dencik, «The assumption is that algorithmic processing of lots of data can serve to anticipate, conjecture and speculate on future behaviour, activities, and threats» (Metcalfe and Dencik, 2019). Ossia l'illusione che la potenza di calcolo dell'IA possa contribuire sia a prevenire attività criminose o semplicemente l'aumento dei migranti irregolari nell'UE sia ad avere un pieno controllo del fenomeno dello spostamento di persone alla ricerca di migliori condizioni di vita. Questo, però, significa fare affidamento sulle capacità predittive dell'analisi di grandi masse di dati informatici, ignorando le soggettività e le esperienze di vita delle persone dalle quali i dati sono estratti. A questo proposito, Lilie Chouliaraki e Myra Georgiou hanno definito la gestione comunitaria delle

frontiere come un «biopolitical regime of border» (Chouraliaki and Georgiou, 2022:9) nel senso foucaultiano dell'espressione, ossia un regime che, a partire dai dati sui singoli corpi e sulle singole identità mira non solo a selezionare chi può entrare e chi no ma anche a produrre conoscenza su chi potrebbe diventare pericoloso, una volta ammesso. Una conferma dell'approccio essenzialmente securitario alla migrazione. Questo rimanda alla vera questione che sta dietro la retorica sugli *smart borders*, quella del potere. Come Ana Beduschi sottolinea, «choice of concepts is not without significance, because it influences the way evidence is created» (Beduschi, 2021:584). Usare strumenti di IA per la gestione delle frontiere vuol dire affidare le decisioni su chi può entrare e a quali condizioni e chi resta fuori ad algoritmi. Algoritmi che sono metodi di calcolo, istruzioni per la rilevazione di regolarità statistiche all'interno di enormi masse di dati al fine di risolvere un problema o per raggiungere uno scopo prefissato. Nel momento in cui questi sistemi imparano e funzionano attraverso i dati che vengono intenzionalmente forniti loro ma che aumentano di continuo con il continuo moltiplicarsi degli ingressi regolari e irregolari, la possibilità di verificarne il funzionamento e la validità dei risultati da parte di operatori umani, come l'*AI act* peraltro richiede, diminuisce drasticamente a prescindere da ogni possibile attenzione in fase di programmazione alla loro trasparenza ed *accountability*. Come gli studiosi sottolineano, chi sarà responsabile di eventuali errori nel riconoscimento o nell'associazione fra dati biometrici che può risultare in un infondato divieto di ingresso (Beduschi 2021; Metcalfe and Dencik 2019)? Ammettendo che i dati che vengono inseriti nei database comunitari siano effettivamente rispondenti ai criteri di assenza di *bias* e discriminazioni di ogni sorta, indicati nell'*IA act*, come ovviare a quelli che lo stesso algoritmo può sviluppare nella sua ricerca di regolarità statistiche fra i dati che lo alimentano (Cristianini 2023)?

Ancora più sottile è la questione sollevata da Dimitri van den Meerssche in un saggio del 2022. Questi sottolinea come la diffusione di sistemi di *smart border* basati sull'estrazione dei dati biometrici dei soggetti sia in realtà una precisa modalità di espressione del potere. Ricorrere a *smart borders* che funzionano a partire dai dati biometrici di coloro che provano ad attraversarli in maniera più o meno regolare vuol dire attribuire la possibilità di ingresso non sulla base di regole certe, stabilite in anticipo e quindi note ma a partire da regolarità statistiche individuate di volta in volta da sistemi di IA che funzionano in modo adattivo rispetto all'obiettivo loro assegnato e ai dati che processano. Se una delle conseguenze è l'indubbia difficoltà nell'individuare il centro di responsabilità per eventuali errori, ancora più importante è la sostanziale deprivazione delle persone alle frontiere. Deprivazione del riconoscimento della loro soggettività ai fini della valutazione delle richieste di protezione internazionale e che sfocia nella loro riduzione in cluster temporanei di dati, smontabili e riaggregabili, per definizione incapaci di ogni soggettività o azione politica (van den Meerssche 2022). Ma fin qui ci muoviamo ancora bene o male all'interno del perimetro delineato dall'*AI act*, anche se ci si potrebbe chiedere quanto il sistema dell'interoperabilità potrà avere effetti analoghi a quelli dei sistemi che lo stesso *AI act* considera ad alto rischio in materia di migrazione. Mi riferisco qui alla possibilità che la verifica dell'identità delle persone, della veridicità delle loro affermazioni o la valutazione dei rischi che il soggetto pone dal punto di vista della sicurezza o della regolarità del suo soggiorno siano nei fatti delegati ai sistemi di IA da operatori di frontiera che si affideranno sempre più ai loro esiti.

4. Verso confini sempre più smart e sempre più chiusi?

Il vero problema è che i dati generano anche bulimia. Più se ne hanno, più se ne vogliono e più cresce la tentazione di moltiplicarne i possibili usi. Attraverso una ricerca certosina fra i progetti finanziati dal programma HORIZON 2000, due ricercatori ne hanno scoperto uno dell'Università di Manchester relativo allo sviluppo di un sistema di IA per il controllo delle frontiere, denominato iBrdrCtrl. Suo punto di partenza è un software di riconoscimento facciale, che è in grado di analizzare i micromovimenti del volto di soggetti sottoposti ad interviste, condotte da un avatar e preliminari al loro ingresso nell'Ue (un possibile complemento di ETIAS?). Sulla base del nesso bugia-colpa-espressione del viso, iBrdrCtrl usa l'analisi dei micromovimenti del viso per stabilire il grado di pericolosità del soggetto intervistato, ossia quantificare le probabilità che intenda prolungare il proprio soggiorno nell'UE diventando un overstayer irregolare o progetti attività illecite. Questo conduce all'attribuzione di un QR code che riflette il grado di pericolosità, stimato dal sistema, del soggetto intervistato e può generare esiti diversi: l'ingresso o il rifiuto di ingresso nell'UE o la richiesta di una nuova e più articolata intervista con un operatore in carne ed ossa (Sánchez-Monedero and Dencik 2022). Testato in precisi varchi doganali dell'Unione, il sistema non ha dato grande prova di sé, come spesso capita quando sistemi di IA passano dalle dimensioni ristrette dei laboratori di progettazione al processare grandi quantità di dati *in vivo* (Jona 2021). Questo, unito alle (poche) interrogazioni a riguardo da parte di europarlamentari, ha fatto sì che il progetto sia stato portato a termine e apparentemente abbandonato. Nei fatti, però, sembra sia stato riformulato e ingrandito. La sua nuova versione si chiama TRESPASS. L'obiettivo è sempre quello di determinare attraverso interviste con un avatar

il livello di propensione dei soggetti che vogliono recarsi nell'UE a prolungare in maniera irregolare il proprio soggiorno o a commettervi attività illegali. A questo fine, all'analisi delle micro-espressioni facciali, che caratterizzava iBorderCtrl, TRESPASS aggiunge la valutazione di altre due serie di dati: da un lato, vengono rilevati gli atteggiamenti/movimenti del corpo degli intervistati durante le interviste. Questi vengono poi analizzati e categorizzati sulla base di modelli di analisi comportamentale. Dall'altro, TRESPASS dovrebbe consentire di confrontare i dati rilevati non solo con quelli già presenti nei database comunitari ma anche con quelli rilevabili su più ampie reti social come Facebook, Twitter, Google+ e Instagram (Ozkul 2023). In questo modo, il sistema dovrebbe consentire di elaborare profili di rischio sempre aggiornati perché costantemente alimentato non solo con i dati rilevati alle frontiere comunitarie ma anche con tutto quello che circola in rete. Un evidente accrescimento di efficienza ma anche di opacità per quanto riguarda l'*accountability* dei suoi esiti, ossia la capacità di spiegare come il suo algoritmo è arrivato all'esito finale. Soprattutto, i sistemi di analisi delle emozioni sono fra quelli che l'*AI act* classifica fra quelli ad alto rischio. E lo *scraping* di immagini facciali, ossia il fatto di ricavarle attraverso processi che non presuppongono il necessario consenso del soggetto all'uso della propria immagine, che è quello a cui potrebbe verosimilmente condurre il set di dati per l'addestramento e il funzionamento di TRESPASS, è vietato in molti Stati membri dell'UE, come hanno ribadito di recente i Garanti della privacy di diversi Stati membri⁸. La Commissione, dunque, finanzia progetti di ricerca che mirano a sviluppare sistemi che le sue norme considerano ad alto rischio e

⁸ Cfr. «Il Garante della privacy olandese multa Clearview AI per l'uso improprio dei dati di riconoscimento facciale» in *Euractive*, 04/09/2024. Nello stesso senso, nel marzo 2022, il Garante italiano della privacy ha sanzionato la stessa società per lo stesso addebito con una multa di 20 milioni di euro. Cfr. <https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9751323>.

soggetti a vincoli in termini di progettazione/distribuzione/uso anche in riferimento al loro impatto sui diritti umani degli individui. O che funzionano sulla base di sistemi che le norme europee ritengono illegali. Un caso di evidente strabismo fra norme e finanziamenti. Ed è difficile pensare che la mano destra non sappia quello che fa la sinistra. Di cosa si tratta, allora? Il punto è che credo siamo davanti ad un caso di pronunciato strabismo. I sistemi di IA stanno plasmando e plasmeranno sempre di più il nostro modo di produrre, lavorare, relazionarci con gli altri, vivere. E non si può che condividere il tentativo dell'Unione europea di orientarne gli sviluppi presenti e futuri in linea con i propri valori di tutela dei diritti individuali. Questi non sono, però, applicati universalmente come amiamo credere (e far credere). Nei fatti, le poco incisive forme di tutela previste dall'*AI act* riguardano i cittadini comunitari. E quelle previste nei regolamenti EES ed ETIAS saranno prevedibilmente attivabili da viaggiatori regolari e consapevoli del percorso burocratico che devono affrontare per entrare nell'UE. Per tutti coloro che partono su mezzi di fortuna per sfuggire a guerre e/o miseria, la tutela dei diritti è molto, molto più limitata. In base ai nuovi regolamenti sulla gestione della migrazione, coloro che arrivano ai confini dell'Unione saranno immediatamente sottoposti al rilevamento dei dati anagrafici, biometrici e a controlli sul loro stato di salute e sul grado di pericolosità. Al termine di questa prima fase, sulla base di un formulario che contiene i dati indicati, i migranti verranno ripartiti in tre gruppi: quelli che non hanno diritto ad entrare e vanno rimpatriati, quelli che possono essere sottoposti a controlli più superficiali in quanto provengono da Paesi con bassi tassi di riconoscimento delle istanze di protezione internazionale o hanno deliberatamente mentito sulla propria identità, coloro che possono depositare domande di protezione ma sono obbligati ad attenderne l'esito in strutture individuate dallo Stato di primo approdo

(Regolamento 2024/1348). Proviamo ad immaginare cosa potrebbe accadere (e nei fatti prevedibilmente accadrà, quando il sistema di interoperabilità entrerà in funzione) se, nello svolgimento di queste procedure, le autorità di frontiera si affidassero a strumenti di IA. I tempi previsti per i controlli sono rigidi e ristretti: come potrebbe una guardia di frontiera accorgersi di errori commessi dal sistema? Su quale base potrebbe farli valere? Quale controllo sui loro dati potranno avere persone, stremate da viaggi lunghi e difficilissimi, per le quali negare il rilevamento dei dati biometrici equivale ad un rifiuto di ingresso? Come potranno venire a conoscenza del diritto a proporre ricorso contro decisioni avverse se nelle strutture nelle quali vengono trattenuti non è prevista la sistematica presenza di ong a loro sostegno e i tempi per presentare eventuali ricorsi sono minimi?

È di questi problemi che il videogioco «Papers, please!» rappresenta un'utile illustrazione. Se i migranti sono byte processati da algoritmi, possono essere spezzettati, analizzati, archiviati e respinti senza alcun problema o moto se non di compassione quanto meno di dubbio. Se ad analizzarne i documenti sono persone, queste avranno la possibilità di vederne le storie e il carico di sofferenza e valutarne il peso rispetto a politiche di gestione della migrazione sempre più orientate alla chiusura ed ai respingimenti. Potrebbe essere il ritorno del diritto cosmopolitico di Kant, quando in *Per la pace perpetua* affermava che poiché la terra è tonda, siamo destinati ad incontrarci. Trovo drammatico che in un mondo nel quale i conflitti armati alle porte dell'Unione si moltiplicano e con loro i profughi, larga parte dei dibattiti fra gli Stati membri ruoti intorno ai modi per chiudere le frontiere o per spostare il carico della gestione dei migranti e delle loro domande di protezione al di fuori del territorio degli Stati membri. «Papers, please!» può servire a capire che decidere chi

entra e chi resta fuori è una responsabilità. Serve anche a me, studiosa a volte disincantata, a ricordare il senso del mio lavoro.

Bibliografia

Beduschi, A. (2021). International Migration Management in the Age of Artificial Intelligence. *Migration Studies*, IX(3), 576-596.

Chouraliaki, L., Georgiou M. (2022). *The Digital Border. Migration, Technology, Power*. New York: New York University Press.

Cirone, E. (2024). L'AI act e l'obiettivo (mancato?) di promuovere uno standard globale per la tutela dei diritti fondamentali. *Quaderni AISDUE*, 2/2024. 1-20.

Commissione europea (2016). Comunicazione della Commissione al Parlamento europeo e al Consiglio. Sistemi d'informazione più solidi e intelligenti per le frontiere e la sicurezza. COM(2016)205 final, Bruxelles, 6.4.2016.

Commissione europea (2021). Proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'unione, COM(2021) 206 final, Bruxelles, 21.4.2021.

Contissa G., Galli F. (2024). AI Act e diritti fondamentali: presupposti tecnologici e ricadute normative. *Quaderni costituzionali*, 3, 738-741.

Cristianini, N. (2023). *La scorciatoia. Come le macchine sono diventate intelligenti senza pensare in modo umano*. Bologna: il Mulino.

European Commission (2018). Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, Artificial Intelligence for Europe, COM(2018)237 final, Bruxelles, 25.4.2018.

European Commission (2019). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Building Trust in Human-Centric Artificial Intelligence, COM(2019)168 final, Bruxelles, 8.4.2019.

European Commission (2020). White Paper On Artificial Intelligence. A European approach to excellence and trust, COM(2020) 65 final, Bruxelles, 19.2.2020.

Finocchiaro, G. (2023). The regulation of artificial intelligence. *AI & Society*, 39,1961–1968. <http://doi.org/10.1007/s00146-023-01650-z> [ultimo accesso: 07.10.2024].

Floridi, L. (2021). «The European Legislation on AI: a Brief Analysis of its Philosophical Approach». *Philosophy & Technology*, 34, 215–222. <http://doi.org/10.1007/s13347-021-00460-9> [ultimo accesso: 07.10.2024].

Glouftsiou, G. (2021). Governing border security infrastructures: Maintaining large-scale information systems. *Security Dialogue*, 52(5), 452–470.

Jona, L. (2021). La “macchina della verità” alle frontiere di cui l’Europa preferiva non parlare. *Wired*, 27.04.2021.

Leese, M. (2020). Fixing State Vision: Interoperability, Biometrics and Identity Management in the EU. *Geopolitics*, 27(1), 113-133.

Lemberg-Pedersen, M. (2024). Crafting markets of exclusion: Frontex and the border security sector. *Mondi migranti*, 2, 9-36.

Lemberg-Pedersen, M., Haioty, M. (2020). Reassembling the Surveillable Refugee Body in the Era of Data-Craving. *Citizenship Studies*, XXIV(5), 607-624.

Metcalfe, P., Dencik, L. (2019). The politics of big borders: Data (in)justice and the governance of refugees <https://firstmonday.org/ojs/index.php/fm/article/view/9934> [ultimo accesso: 27.10.2024].

Novelli, C. *et al.* (2023). «Taking AI risks seriously: a new assessment model for the AI Act». *AI & Society*, 38(3). <http://doi.org/10.1007/s00146-023-01723-z> [ultimo accesso: 03.10.2024].

Ozkul, D. (2023). *Automating Immigration and Asylum. The Uses of New Technologies in Migration and Asylum Governance in Europe*. Refugee Studies Centre, University of Oxford.

PICUM (2019). Data Protection, Immigration Enforcement and Fundamental Rights: What the EU’s Regulations on Interoperability Mean for People with Irregular Status su <https://picum.org/wp-content/uploads/2019/11/Data->

Protection-Immigration-Enforcement-and-Fundamental-Rights-Full-Report-EN.pdf [ultimo accesso: 15.10.2024].

Regolamento (UE) n. 1077/2011 del Parlamento europeo e del Consiglio, del 25 ottobre 2011, che istituisce un'agenzia europea per la gestione operativa dei sistemi IT su larga scala nello spazio di libertà, sicurezza e giustizia. GUUE L 286, 1.11.2011, 1-17.

Regolamento (UE) 2019/817 del Parlamento europeo e del Consiglio del 20 maggio 2019 che istituisce un quadro per l'interoperabilità tra i sistemi di informazione dell'UE nel settore delle frontiere e dei visti e che modifica i regolamenti (CE) n. 767/2008, (UE) n. 2016/399, (UE) 2017/2226, (UE) 2018/1240, (UE) 2018/1726 e (UE) 2018/1861 del Parlamento europeo e del Consiglio e le decisioni 2004/512/CE e 2008/633/GAI del Consiglio in GUUE L 135 del 22.5.2019, pp. 27-84.

Regolamento (UE) 2024/1348 del Parlamento europeo e del Consiglio del 14 maggio 2024 che stabilisce una procedura comune di protezione internazionale nell'Unione e abroga la direttiva 2013/32/UE https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=OJ:L_202401348.

Regolamento (UE) 2024/1358 del Parlamento europeo e del Consiglio del 14 maggio 2024 che istituisce l'«Eurodac» per il confronto dei dati biometrici ai fini dell'applicazione efficace dei regolamenti (UE) 2024/1351 e (UE) 2024/1350 o del Parlamento europeo e del Consiglio e della direttiva 2001/55/CE del Consiglio e ai fini dell'identificazione dei cittadini di paesi terzi e apolidi il cui soggiorno è irregolare, e per le richieste di confronto con i dati Eurodac presentate dalle autorità di contrasto degli Stati membri e da Europol a fini di contrasto, che modifica i regolamenti (UE)

2018/1240 e (UE) 2019/818 del Parlamento europeo e del Consiglio e che abroga il regolamento (UE) n. 603/2013 del Parlamento europeo e del Consiglio <http://data.europa.eu/eli/reg/2024/1358/oj> [ultimo accesso: 14.10.2024].

Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale) <http://data.europa.eu/eli/reg/2024/1689/oj> [ultimo accesso: 28.10.2024]

Rida Nour, M. (2019). Geopolitique de l'intelligence artificielle: Les enjeux de la rivalité sino-américaine. *Paix et Sécurité internationale*, 7, 231-259.

Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica.

https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_IT.html#title1 [ultimo accesso: 17.10.2024].

Sánchez-Monedero, J., Dencik, L. (2022). The Politics of Deceptive Borders: 'Biomarkers of Deceit' and the Case of iBorderCtrl. *Information, Communication & Society*, XXV(3), 413-430.

Tazzioli, M. (2022). Extract, datafy and disrupt. Refugees' subjectivities between data abundance and data disregard. *Geopolitics*, 27(1), 70-88.

Van Den Meerssche, D. (2022). Virtual Borders: International Law and the Elusive Inequalities of Algorithmic Associations. *European Journal of International Law*, XXXIII(1), 171-204.

Von der Leyen, U. (2024). La scelta dell'Europa. Orientamenti politici per la prossima Commissione europea 2024-2029. https://commission.europa.eu/document/download/e6cd4328-673c-4e7a-8683-f63ffb2cf648_it?filename=Political%20Guidelines%202024-2029_IT.pdf#:~:text=Gli%20orientamenti%20politici%20sono%20il%20nostro%20piano%20per%20la%20forza [ultimo accesso:30.09.2024].

8. SOVRANITÀ E INDIPENDENZA TECNOLOGICA. LA SFIDA E I RISCHI DELLE “NUOVE” INTELLIGENZE. UNA VALUTAZIONE DI SISTEMA

di Giuseppe Romeo*

Abstract: Non vi sono dubbi che l’approccio verso una “nuova” dimensione di intelligenza diretta ad aggiungersi alle possibilità umane – nel tentativo di amplificarne le capacità e le possibilità di elaborazione di comportamenti, oltre che di informazioni – comprenderà e coinvolgerà la sfera esistenziale dell’individuo nella sua dimensione di persona quanto nelle sue manifestazioni organizzate di soggetto politico ed economico. Tutto questo, nelle diverse e nuove architetture organizzative destinate a gestire, quindi a governare ogni futuro della comunità internazionale e, con essa, dell’intera umanità rideterminando nuovi rapporti di forza e nuove relazioni tra soggetti politici internazionali. La sfida al dominio delle superintelligenze, nel campo dell’IA come nel Quantum-Computing determinerà nuovi assetti e nuove modalità di governance cui dovranno corrispondere nuove norme e nuove strategie rivolte a non fare della contrattazione internazionale una pericolosa occasione per politiche di potenza indirizzate alla conquista della supremazia tecnologica in un modello di sistema internazionale tecnopolare.

Parole chiave: colonialismo tecnologico, governance tecnologica, indipendenza e sovranità tecnologica, supremazia tecnologica, tecnopolarità.

* Scuola Universitaria Interdipartimentale in Scienze Strategiche, Università degli Studi di Torino. [manca mail](#)

Accettato Dicembre 2024 - Pubblicato Aprile 2025.

Abstract: There is no doubt that the approach towards a “new” dimension of intelligence aimed at adding to human possibilities – in an attempt to amplify the capabilities and possibilities of processing behaviour, as well as information – will include and involve the existential sphere of the individual in his dimension as a person as well as in his organized manifestations as a political and economic subject. All this, in the different and new organizational architectures destined to manage, and therefore to govern, every future of the international community and, with it, of all humanity, redetermining new balances of power and new relationships between international political subjects. The challenge to the dominance of superintelligences, in the field of AI as in Quantum Computing, will determine new structures and new modes of governance which will have to correspond to new rules and new strategies aimed at not making international bargaining a dangerous opportunity for power policies aimed at conquering technological supremacy in a model of technopolar international system.

Keywords: technological colonialism, technological governance, technological independence & sovereignty, technological supremacy, technopolarity.

1. Nuove traiettorie del sistema-mondo. Una premessa

A volte si prova a immaginare un mondo nuovo nei suoi contorni non solo culturali ma, in particolare, economici, politici e giuridici, dove il senso dell’esperienza umana, nelle sue diversità manifestate nelle varie declinazioni geografiche e geopolitiche si assumono a vere e proprie nuove frontiere della lotta per il potere inteso, quest’ultimo, come raggiungimento di una nuova egemonia. Ogni sforzo nella corsa al dominio tecnologico sottende, infatti, la necessità di garantirsi per ogni attore, politicamente ed economicamente organizzato, la propria sopravvivenza fisica e politica quale parte di un modello di relazioni dettato dalla forza intesa,

quest'ultima, sempre di più in una dimensione ibrida; ovvero, non convenzionale, seppur sempre manifestazione di una capacità di coartare volontà altrui per ricondurne i destini di Stati e individui verso i desideri e gli interessi dell'attore dominante. Che si tratti di potenze militari o di potenze economiche poco importa, dal momento che ogni conquista tecnologica è parte di un nuovo modo di percepire il sé geopolitico e la possibilità di conquistare spazi di potere lasciando spesso al diritto un ruolo ordinatore, certo, ma solo se condiviso dal più forte con buona pace delle migliori tradizioni del liberalismo classico e di ciò che resta del cosmopolitismo kantiano. È, quindi, evidente, che la sfida non è solo politica e geo-politica, ma anche "geo-giuridica" dal momento che proprio lo stabilire regole internazionali che possano ricondurre l'uso delle nuove intelligenze all'interno di un quadro normativo condiviso dovrebbe permettere, e garantire, una transizione verso un modello auspicabilmente tecno-cooperativo delle relazioni internazionali. D'altronde, in questa nuova dimensione geopolitica che vede il potere tecnologico essere sempre più centrale nel determinare le relazioni internazionali in tutte le loro manifestazioni – politiche, economiche e culturali e, soprattutto, giuridiche laddove il diritto possa ancora avere un senso al di fuori della sfera degli ordinamenti nazionali – si risolverà la vera sfida e si determineranno i veri significati attribuibili a istituti come la stessa sovranità. Quest'ultima, espressa in un quadro di un'indipendenza tecnologica e di condizioni di accesso a pari opportunità di utilizzo delle risorse digitali nelle loro diverse destinazioni d'uso evitando il rischio di pericolose asimmetrie. Aspirazioni che, se non tutelate da regole certe e da un'etica consapevole, troveranno il loro contrario nella colonizzazione algoritmica delle possibilità di scelta, nel subire le iniziative altrui. Un pericolo che vedrebbe l'attore meno capace subordinare il proprio agire alla necessità di garantirsi un possibile

standard di vita accettabile, seppur rinunciando a essere protagonista del proprio futuro, quindi rinunciando alla propria libertà. In questa prospettiva, si disegna una nuova singolarità dettata da dipendenze da forme evolute di IA nell'ambito dei processi di decision-making e di governance ad ampio spettro dove, in assenza di un ordine certo di accesso alla tecnologia e alle sue risorse, le prossime traiettorie del «sistema-mondo» lasceranno il campo a transizioni continue, ad ambiguità etiche e giuridiche e a rideterminazioni tra nuovi centri e nuove periferie digitali per e nelle quali chi dominerà la sfida tecnologica, e le risorse per essa necessarie, dominerà i destini dell'umanità al di là di ogni disciplina giuridica possibile¹. Trovare punti di non ritorno nella storia dell'umanità non è certo difficile e non è, altrettanto, *rara avis* comprendere che la progressione tecnologica del supporto al quotidiano rappresenti la rotta verso la quale si estrinseca non solo la semplificazione dei processi più elementari dell'esistenza umana ma, anche, tutto ciò che rappresenta termini e modi per attribuire all'esistenza stessa un significato non solo di scoperta ma di relazione con l'altro, sia esso espresso in termini cooperativi che competitivi. Non vi sono dubbi, quindi, che l'approccio verso una “nuova” dimensione di intelligenza diretta ad aggiungersi alle possibilità umane – nel tentativo di amplificarne le capacità e le possibilità di elaborazione di comportamenti, oltre che di informazioni – comprenda e coinvolga la sfera esistenziale dell'individuo nella sua dimensione di persona e nelle sue

¹ Circa la teoria del «sistema-mondo» è Immanuel Wallerstein che – nel descrivere il «sistema-mondo» come una realtà sociale composta da nazioni, aziende, famiglie, classi e gruppi identitari di ogni tipo interconnessi – identificava ed evidenziava il significato dei momenti chiave nell'evoluzione del moderno sistema-mondo. Ovvero, con lo sviluppo di un'economia-mondo capitalista nel XVI secolo, l'inizio di due secoli di centrismo liberale nella Rivoluzione francese del 1789, e l'indebolimento di quel centrismo nelle rivolte globali del 1968 (Wallerstein 2004).

manifestazioni organizzate quali parti di architetture umanizzate rivolte a gestire, quindi a governare, ogni scelta e ogni futuro.

In un'ottica di massimi sistemi, che impegnerà ruoli e funzioni delle categorie politiche più tradizionali attraverso le quali si organizzano le relazioni umane – Stati, organizzazioni sovranazionali, alleanze – un approccio politologico sull'impatto dell'IA in termini di nuovi scenari geopolitici, ovvero di sistema sembra, quindi, essere non solo necessario, ma non rinviabile.

2 Tecnopolarità. Il paradigma del presente che dominerà il futuro

Le transizioni digitali, come sperimentato in questi ultimi anni, trasformeranno sempre di più l'universo dell'umana quotidianità in ogni sua espressione: economica, sociale, culturale se non politica e strategica. Intelligenza artificiale, blockchain, l'Internet delle cose (*Internet of Things* - IoT) e la robotica apriranno nuove prospettive di convivenza, di determinazione di nuovi rapporti di forza e assumeranno il ruolo di definire nuove modalità comportamentali. Per questo, è evidente che settori come le nuove tecnologie e i nuovi processi di elaborazione delle informazioni, quanto le applicazioni dell'IA non potranno che porre interrogativi e richiedere soluzioni che dovranno sottendere nuove morfologie delle società umane e della comunità internazionale nella sua espressione più ampia. Ovvero, ridefinendo non solo modi e termini di difesa e garanzia di diritti ritenuti indiscutibili e, quindi, inviolabili come i diritti umani universalmente accettati dalle convenzioni e statuti oggi in vigore. Quanto nel definire regole necessarie per l'impatto che l'IA determinerà sulla convivenza e i termini di relazione tra potere e individuo e tra gli individui (Peyron 2021). Il rischio, infatti, è una nuova polarizzazione del «sistema-

mondo» non solo in termini di ricchezza, ma di tecnologia disponibile e di accesso che potrà determinare nuove gerarchie tra i detentori delle migliori e più performanti conoscenze nel campo della IA e delle risorse necessarie e chi di fatto ne sarebbe escluso se non ricondotto a termini di dipendenza. In questa nuova prospettiva delle relazioni internazionali, ci dovremmo abituare a confrontarci con nuove visioni e percezioni di un mondo che tenderà a spostarsi su una iperglobalizzazione dettata dalla necessità di ricercare pari opportunità di competizione attraverso il dominio delle risorse necessarie e delle conoscenze tecnologiche. Tutto questo, contrassegna già oggi quella «tecnopolarità» che si delinea quale ordine mondiale emergente in cui operano le più grandi aziende tecnologiche e Stati nazionali rivali come attori principali negli affari internazionali. Ian Bremmer, a tal proposito, definisce chiaramente un modello di relazione tecnopolare la quale si differenzierà dal passato perché la sovranità e l'influenza non sono determinate dal territorio fisico e dalla sola potenza militare, ma dal controllo su dati, algoritmi e server ((Eurasia Group 2023).

Bremmer ha espresso molto chiaramente il concetto di tecnopolarità laddove chiarisce come e in che misura la rivoluzione digitale, e si può aggiungere l'era delle nuove intelligenze, ha di fatto favorito la transizione da un modello di relazioni internazionali con gli Stati-nazione al vertice della piramide degli attori geopolitici a un ordine tecnopolare aperto a nuovi interpreti delle relazioni economiche, politiche e... tecnologiche dei prossimi anni (Bremmer 2021). Per Bremmer, il «sistema tecnopolare» si fonda su due vantaggi fondamentali offerti dalle nuove tecnologie. Il primo, è rappresentato dal fatto che le aziende tecnologiche, rispetto agli Stati nazionali, dominano lo spazio digitale. Una dimensione, oltre che uno spazio, completamente nuovo che determina una distribuzione dei termini di potenza e, quindi, geopolitici all'interno e per gli effetti dei quali sempre più persone

vivranno la propria esistenza, il proprio quotidiano e i cui bisogni si risolveranno nelle modalità di accesso a un universo virtuale di offerta di servizi. °Un universo di opportunità governate dal remoto delle proprie vite e di cui lo stesso metaverso, ad esempio, ne rappresenta l'architettura più immediata a oggi. Insomma, uno dei risultati più immediati è che ciò permetterà ai detentori del dominio tecnologico di determinare le opportunità e le aspettative dei cittadini degli stessi Stati in un mondo governato da attori capaci di penetrare l'intimo digitale attraverso la creazione di modelli paralleli di relazione. Il secondo, è rappresentato dal fatto che gli attori tecnologici, quali detentori di una quota-parte della capacità di agire nel sistema internazionale, saranno gli unici a fornire una quota sempre più crescente delle capacità/opportunità digitali già evolute, disponendo della proprietà delle infrastrutture fisiche necessarie per far funzionare, in modo moderno, economie e società: dagli smartphone, alle telecomunicazioni, alle reti logistiche, al *cloud* dei servizi, alle piattaforme di e-commerce, ai sistemi di pagamento. Ciò attribuisce loro il controllo su quelle tecnologie simili all'Intelligenza Artificiale che guideranno la prossima rivoluzione tecno-industriale rideterminando i termini stessi di relazione tra gli attori internazionali operando, se non sovrapponendosi negli stessi ambiti in cui operano i governi. Tuttavia, per Bremmer, l'influenza geopolitica delle aziende tecnologiche non potrà che crescere considerati i più ampi e diffusi spostamenti già oggi consolidatisi nel quotidiano digitale di ogni individuo, con il pericolo di dare forma a formule di governance diffusa e penetrativa. Formule, queste ultime, tali da rendere marginali le azioni dei governi il cui monopolio del potere di direzione delle dinamiche politico-economiche e politico-sociali rischia di essere determinato in un altrove geopolitico. Una periferia virtuale che vede la presenza di nuovi attori capaci, ad esempio, di intervenire sul consenso politico dell'opinione pubblica

definito ormai su piattaforme social con la possibilità di “deplatformare” (*deplatform*), ovvero di escludere profili e informazioni provenienti da leader politici o di promuoverli (Van Dijck, de Winkel & Schäfer, 2021). È evidente, insomma, che ciò che la tecnopolarità richiede è una presa di coscienza dei vantaggi ma anche dei pericoli che derivano dall’iperconnessione dei processi di governance quanto dall’iperautomazione degli stessi. Tutto questo, in un modello relazionale che vede coniugarsi le necessità di definire nuove politiche di condotta, oltre che sistemi normativi adeguati, che ridefiniscano termini e modi di accesso e di uso di sistemi che vedono dimensione cyber e nuove intelligenze presentarsi come realtà complementari, perché necessariamente tecnologicamente interdipendenti. In questa prospettiva, la tecnopolarità richiede, allora, un cambiamento di paradigma su come e in che termini andrebbero ridisegnate e reinterpretate le relazioni internazionali e, quindi, la stessa nuova governance globale.

3. Supremazia tecnologica

Ogni competizione tecnologica, come quella che si ravvisa nel confronto tra Cina e Stati Uniti, viene ricondotta ancora oggi nell’alveo di un paradigma statalista descrivendo, riportando la riflessione di Bremmer, secondo il quale le aziende tecnologiche costituiscono soldati di fanteria in un conflitto tra paesi ostili (Bremmer, 2021). Ma le aziende tecnologiche non sono semplici strumenti nelle mani dei governi (Bremmer, 2021). Nessuna delle loro azioni avviene per volere dei governi o degli apparati di sicurezza, risolvendosi in decisioni private prese da società che di certo non perseguono necessariamente finalità filantropiche e che esercitano il loro potere sui codici, sui server e sulle normative

sotto il loro controllo. Il nuovo «sistema-mondo», che ricolloca al centro del futuro prossimo l'Asia per differenza di capacità tecnologiche non raggiunte dall'Europa, richiede la costruzione di un equilibrio di potere determinato tanto dalla competizione tra Stati-nazione e aziende tecnologiche quanto dalla concorrenza tra Stati Uniti e Cina. Attori geopolitici, questi ultimi, che nello spostare la competizione verso l'egemonia globale sul campo del differenziale tecnologico proveranno a contenere – riconducendo al proprio interno attori non-Stato tecnologicamente ed economicamente determinanti – l'inadeguatezza attuale del paradigma post-westfaliano che tende a far sopravvivere un modello di equilibrio garantito dagli Stati-nazione non più adeguato per interpretare un futuro a n-attori. È, quindi, chiaro, il perché, una regolamentazione efficace delle tecnologie emergenti come l'Intelligenza Artificiale richiederà l'assegnazione di un ruolo formale alle principali aziende tecnologiche insieme ai governi. In questo senso, può essere letta la volontà del neo-presidente Donald Trump di inserire nel nuovo board della sua Presidenza, Elon Musk patron di Tesla. Ciò, ovviamente, comporterebbe anche la costruzione se non la promozione di nuovi profili etici e morali. Profili, che non necessariamente saranno percepiti come uguali al passato. Ma che si determineranno adattandosi alle nuove possibilità di relazione e di processo offerte da una e-governance e da una partecipazione digitale che presupporrà una nuova visione, e versione, della responsabilità di chi produrrà nuove norme e chi stabilirà nuove forme di accesso e di utilizzazione alla migliore qualità della tecnologia digitale e IA. In questo senso, ogni aspetto introdotto nell'universo sempre più digitalizzato non potrà che veder man mano affermarsi una nuvola informatica capace di comprendere, nel suo estendersi a ogni ambito della vita di una comunità o di una società, un *cloud* di possibilità di semplificare la vita di ogni cittadino come di esporne le debolezze, le aspettative e

le ansie. D'altronde, universalizzare le funzioni dell'IA, alla stessa stregua di come si riteneva di rendere universale valori etici e morali, rappresenta un rischio o un'opportunità. Il tutto dipenderà dalle nuove regole stabilite dai detentori delle migliori opportunità di governance tecnologica esprimibile. Una condizione necessaria, questa, che sottende anche il superamento del paradigma della linearità del pensiero tecnologico. Un superamento espresso da una nuova possibilità di vedere persistere una volontà di universalizzare percezioni che propongono giudizi di valore e che giustificano comportamenti o azioni politiche tali da favorire nuove egemonie positive, perché condivise e tali da ridurre i rischi di un universalismo tecnocratico incapace di universalizzare i particolari di ognuno quanto di particolarizzarne gli universali (Wallerstein, 2007: 63-64). Con tali premesse non è certo un azzardo di stile affermare che, come nel passato, la tecnologia non è sicuramente un campo neutrale alle dinamiche geopolitiche ma neanche neutrale alle implicazioni sociali. Essa, si presenta nella storia dell'umanità come un acceleratore evolutivo quale espressione del superamento progressivo di conoscenze che, nelle sempre più complesse modalità di organizzazione del potere, si assume il ruolo di definire le possibilità di affermazione di egemonie costruite sulle migliori disponibilità di nuove conoscenze tecniche a supporto del confronto politico, economico come di quello militare. D'altra parte, se la supremazia tecnologica era ritenuta quale fattore di potenza nella corsa alla supremazia nucleare, allo stesso modo essa si ripropone oggi quale condizione ancora più efficace per imporre una *supremacy* tecnologica nel campo delle capacità di dominio in una realtà relazionale a n-dimensioni. Questo, traghettando opportunità dal reale analogico alla dimensione cyber-quantistica e delle nuove intelligenze, assumendo un livello di penetrabilità delle comunità mai raggiunto sino ad oggi. In questo senso, parole come resilienza, sovranità tecnologica leadership diventano

complementari in ogni ambito relazionale poiché, ad esempio, non esisterà leadership senza una capacità di dominare assetti e processi di decisione e di azione che siano supportati da competenze tecnologiche tali da fare la differenza con i propri competitor. Ciò significa, che in molti Stati occidentali le relazioni con le Big Tech non sempre saranno idilliache poiché la sovranità digitale non sempre sarà interpretata come un concetto giuridico costituzionalmente qualificato e/o come un valore “politico” condiviso. Insomma, il differenziale tecnologico rappresenterà un buon argomento, se non un fattore decisivo, per definire la transitorietà degli stessi schieramenti possibili e configurabili di un «sistema-mondo» che sarà funzione della migliore tecnologia esprimibile in termini di produzione o, comunque, anche della migliore e più conveniente accessibilità. E, se nel passato si poteva ritenere, con giusta osservazione, che l'equilibrio di potenza si adattava alla distribuzione globale delle risorse economiche oggi si potrebbe affermare che l'equilibrio di potenza dipenderà, o sarà funzione, della distribuzione globale, e d'eguale, dell'accesso alla tecnologia e alle risorse che supportano i modelli tecnologici (Kennedy, 2001: 25). Ciò significa che, superati i termini classici della competizione ancorata alle sole capacità militari e alle possibilità finanziarie, assumere una posizione di leadership internazionale risiederà nella forza di ridisegnare le stesse geografie di ieri ma dove i confini del passato saranno superati da nuove opportunità, se non di competizione, nei settori sempre più vitali del quotidiano di ogni attore, sia esso Stato o individuo. Ma non solo. La nuova geografia, che sarà determinata dalle relazioni costruite dagli Stati o disegnate dalle stesse Big Tech, farà sì che la nuova carta del mondo vedrà man mano spostarsi l'idea di centro con il decentramento delle comunità e riconfigurazione dell'idea stessa di periferia come luogo-limite. Spazio, quest'ultimo, dove si risolveranno i rapporti di potenza a scapito degli Stati che saranno

costretti a rivedere non solo gli assetti di governance interna, ma anche di valutare se, come e in che misura riorganizzare, con sempre maggiori concessioni di autonomia, le comunità che li compongono. Tutto questo, amplificato dalla corsa a usufruire delle migliori connessioni con i mercati che contano e dalla concentrazione del potere finanziario in luoghi nei quali si risolve la libera determinazione del valore della ricchezza. D'altronde, la stessa rinascita delle città-Stato, ad esempio, così come ipotizzata da Parag Khanna, dimostra quanto l'aspetto dimensionale non abbia più un carattere decisivo nell'imporsi con i suoi pesi demografici ed economici, dal momento che è la capacità di poter gestire i destini tecnologici a fare la differenza nel domani prossimo. Un'abilità di gestire la ricchezza mondiale all'interno di spazi autodeterminati e a forte connessione per i quali finanza e geografia sino a ieri dettavano, in sostanza, le regole della competizione. Oggi, finanza e connessioni possibili spostano in avanti il livello di competizione (Khanna 2017a, 2017b). Ma non solo. Anche quelli che negli anni Novanta sembravano essere approcci innovativi nell'interpretare le relazioni internazionali, andando a prendere in prestito strumenti e categorie tipiche della sociologia, trovano una nuova prospettiva laddove, ad esempio, anche il superamento della relazione centro/periferia rappresenta un fattore decisivo per guardare a un sistema multicentrico. A queste condizioni, non vi è dubbio, infatti, che la stessa IA rideterminerà il rapporto tra centro e periferia soprattutto con una nuova ricollocazione dei processi produttivi. Infatti, se il «sistema-mondo» preindustriale è stato demograficamente rurale, oggi la ruralità si è persa in una concezione definita da tanti centri ognuno dei quali presenta una sua dignità tecnologica, facendo sì che il concetto stesso di periferia si riduca a un altrove remoto, ma non per questo da intendersi ai limiti delle possibilità di interagire con altri attori. Un sistema, dove l'idea dell'esistenza di una

semiperiferia cerca di rideterminare le relazioni tra paesi che hanno subito una rapida crescita economica attraverso l'industrializzazione orientata, quest'ultima, all'esportazione dalla Seconda guerra mondiale in poi, esprimibile secondo tre approcci diversi, ma interconnessi, utilizzati sino alla fine del Novecento per spiegare l'ascesa delle Nuove Economie Internazionali (NIE- Newly Industrializing Economies).

Per Wallerstein, le tre chiavi interpretative erano da individuare in: un approccio del mercato liberale (commercio internazionale e sugli investimenti transfrontalieri); in un approccio dello stato evolutivo (ruolo degli interventi statali dietro lo sviluppo); in un approccio di aggiornamento tecnologico (mezzi con cui le NIE recuperano il ritardo tecnologico) (Wallerstein 1997)² Riletto oggi, si potrebbe dire che, al netto delle versioni neoliberaliste e i prodotti della globalizzazione quali risultati di un modello economico fondato sul libero mercato, ma non per tutti, il terzo approccio sembra prevalere soprattutto se ripositionato sul piano della competizione tecnologica nel campo non solo della declinazione digitale, ma delle nuove intelligenze e delle opportunità offerte dalla meccanica quantistica. D'altra parte, nelle previsioni e garanzie e nelle diverse possibilità di fissare delle regole vi è racchiusa un'ambiziosa volontà diretta a proteggere la sovranità digitale di ogni *player* internazionale – sia esso Stato o attore appartenente alle NIE – sfruttando strumenti e poteri normativi per definire regole e standard globali in un un momento

² Riferito a un momento storico della fine del Novecento, Wallerstein riteneva che i nuovi cambiamenti dovuti dalla fine della Guerra Fredda e da un processo di presunta globalizzazione dei termini di relazione, siano essi politici che economici, avrebbero visto determinarsi una congiuntura di tre grandi punti di svolta, ciascuno dei quali portatore di un fenomeno recessivo. Questi comprendevano, nel breve termine, la fine dell'espansione economica mondiale nell'attuale ciclo di Kondratieff di 50 anni; nel medio termine, l'inizio del declino del ruolo dominante degli USA; e, a lungo termine, segnali di una possibile disintegrazione sistemica dell'economia-mondo capitalista vecchia di 500 anni.

di forte corsa verso la supremazia tecnologica, come dimostrano Cina e Stati Uniti ad esempio.

4. Nuove egemonie

Nella nuova sfida tecnologica l'idoneità di una società, la capacità di un sistema sociale di far fronte a sfide e opportunità complesse come quelle che saranno proprie della nuova era dipenderà ancora una volta dai valori e dallo stile di vita che ogni Stato, ogni società e ogni realtà politicamente se non economicamente organizzata guadagnerà quali posizioni di supremazia, mentre altri/e li/le perderanno³. Gli stessi studi sulla complessità,

³ La «teoria della complessità» è ancorata a dei concetti di base quali: 1. «Idoneità sociale» (*fitness*): capacità di far fronte alla complessità. Per sopravvivere alle sfide e sfruttare al meglio le opportunità, un organismo in forma può elaborare informazioni e gestire molte variabili; 2. «Coevoluzione» (*coevolution*): ogni individuo, specie e società coevolve con gli altri e con il loro ambiente condiviso. Un cambiamento in un qualsiasi attore o ambiente può alterare l'idoneità di più attori. Più variabili modellano un sistema, più difficile prevedere come il cambiamento in un elemento influenzerà gli altri (l'«effetto farfalla»). La non linearità e la complessità sono tratti distintivi delle reti sociali umane. La teoria della complessità cerca di spiegare il processo di adattamento complesso all'interno di sistemi complessi, siano essi ecosistemi, Internet o sistemi politici; 3. «Emergenza»: la coevoluzione spesso dà origine a «proprietà emergenti»; 4. «Sistemi basati su agenti»: un sistema basato su agenti è quello in cui attori indipendenti, ciascuno seguendo alcune regole, si auto-organizzano per formare un fenomeno emergente senza una direzione centrale dall'alto. Senza pianificazione, cooperano per il reciproco vantaggio. Come una barriera corallina, ogni ecosistema durevole è un fenomeno emergente; 5. «Auto-organizzazione»: la chiave del *fitness* è l'auto-organizzazione risultante dalle azioni, pianificate o meno, da agenti indipendenti. Insieme, generano un sistema che affronta bene la complessità; 6. «Criticità auto-organizzata»: in equilibrio tra ordine e caos. La teoria della complessità avverte che le società potrebbero essere meno in forma di quanto sembri. Il *fitness* dipende dall'armonia di molti fattori; 7. «Equilibrio puntiforme»: il concetto di equilibrio puntiforme sottolinea che l'evoluzione è spesso segnata da ondate di speciazione e valanghe di estinzione. Le specie spesso si sviluppano rapidamente, persistono a lungo con pochi cambiamenti e

quale contraltare al determinismo lineare fondato sulla reversibilità del tempo così come interpretato da Isaac Newton piuttosto che dallo stesso Albert Einstein, hanno cercato di dimostrare, secondo Wallerstein, quanto i sistemi non siano lineari, ma tendono a muoversi lontano dalle condizioni di equilibrio (Wallerstein, 2007: 87). La sfida, in altre parole, sarebbe data dal fatto che è intrinsecamente impossibile determinare le future traiettorie di qualsiasi proiezione. Ciò significherebbe, tradotto nei termini del vissuto contemporaneo, che neanche riferendosi all'IA si possono oggi, a conoscenze date, definire le possibili traiettorie evolutive e questo perché scienza e tecnologia non riducono il complesso al semplice, ma tendono a spiegare, la prima, e risolvere, la seconda, livelli di ulteriore complessità seguendo una «freccia del tempo» secondo Wallerstein, che non rende reversibili certi processi evolutivi nei diversi settori dell'esperienza fisica e dell'esperienza sociale (Wallerstein, 2007). Ciò significa, nell'ultimo caso, che se la spinta ad eccellere – come individui e come comunità – dovrebbe essere sufficiente per competere nel nuovo secolo, altrettanto dovrebbe essere anche la capacità di condividere valori comuni attraverso la cooperazione e condivisione delle opportunità offerti dall'IA così come anche i rischi. Si tratta, quindi, di valutare un approccio compatibile nel poter cercare di interpretare le traiettorie di domani nella competizione tecnologica sottesa all'uso delle nuove intelligenze, affidandosi alla scienza della complessità per la quale il «sistema-mondo» vede i suoi protagonisti, Stati e aggregati umani ed economici, come soggetti interdipendenti, reciproca-

poi si estinguono improvvisamente, non gradualmente. Grazie alla mutazione e all'auto-organizzazione, i membri della specie trovano la loro nicchia e si aggrappano ad essa. Quando il loro ambiente cambia, devono adattarsi o scomparire; 8. «Paesaggi dell'idoneità» (o del *fitness*): la teoria della complessità suggerisce che la coevoluzione può essere mappata come un paesaggio aspro in cui l'idoneità relativa di ciascun organismo è mostrata come un picco che sale o scende come conseguenza della coevoluzione. A tal proposito cfr., Clemens (2013).

mente vulnerabili e, soprattutto, così strettamente legati da potersi aiutare così come danneggiarsi a vicenda. A tal proposito, anche il concetto di idoneità sociale, una delle principali preoccupazioni della «teoria della complessità», comprende propri punti di forza politici, economici e culturali. Il ruolo preciso svolto da ciascuna forza nel plasmare le migliori prestazioni sociali diventa una questione importante, ma dovremmo aggiungere anche aspetti tecnologici determinanti i nuovi rapporti di forza in senso lato⁴. È evidente, allora, che l'ingresso di nuove possibilità tecnologiche offerte alle relazioni tra Stati e attori politicamente ed economicamente organizzati, fa sì che la complessità diventi un aspetto caratterizzante i processi e le interazioni politiche globali, ed è stato così negli ultimi tre decenni, delineando i vantaggi dell'integrazione di questa teoria nella disciplina delle relazioni internazionali. Il presupposto è che un sistema internazionale debba essere interpretato come un sistema complesso e disordinato (o un sistema complesso costituito da molti sistemi complessi), piuttosto che dall'ipotesi tradizionale che lo presenta come un sistema con una struttura caotica (Hemchi, 2022)⁵.

⁴ La «teoria della complessità» enfatizza le interazioni e i relativi cicli di feedback che cambiano costantemente i sistemi. Sebbene si proponga che i sistemi siano imprevedibili, essi sono anche vincolati da regole che generano ordine. Le pubbliche relazioni internazionali si basano sulle relazioni umane. Le relazioni umane implicano una complessa interazione di emozioni, impressioni e motivazioni. La complessità delle relazioni internazionali non dovrebbe quindi sorprenderci. Ciò è particolarmente vero quando si comprende che le relazioni tra le nazioni devono tenere conto della somma totale delle pressioni e degli stress che colpiscono intere popolazioni (Clemens, 2013 e anche Argyle, Kohara, 1985).

⁵ In quest'opera Hemchi sostiene che la complessità è un aspetto essenziale dei processi e delle interazioni politiche globali, ed è stato così negli ultimi tre decenni, delineando i vantaggi dell'integrazione di questa teoria nella disciplina delle relazioni internazionali. Si parte da una comprensione del sistema internazionale come un sistema complesso e disordinato (o un sistema complesso costituito da molti sistemi complessi), piuttosto che dall'ipotesi tradizionale che lo presenta come un sistema con una struttura caotica, che è stata mantenuta dalle successive relazioni internazionali. Qui, Hemchi apre una discussione su diversi

Tuttavia, complessità o meno, cercando una ragione di ordine necessario senza perdere di vista i nuovi altrove tecnologici, anche un attento Henry Kissinger non ha lesinato di definire la sfida tecnologica nel campo delle nuove intelligenze come la sfida per la sopravvivenza di un'umanità che dovrà fare i conti con nuovi paradigmi che richiederanno la capacità di esprimere una nuova forma di governance internazionale. Una governance, che inciderà decisamente non solo nell'indicare come e in che misura si potranno definire nuove politiche di potenza, quanto il comprendere il ruolo che l'IA assumerà nel riorganizzare le stesse relazioni umane nei diversi settori dell'esistenza, dal punto di vista politico, economico, sociale e tecnologico. In *The Age of AI: and Our Human Future* (Kissinger, Schmidt, Huttenlocher, 2023) si sottolinea, infatti, come l'Intelligenza Artificiale abbia non solo imparato a vincere a scacchi superando nelle strategie di gara mosse che grandi campioni non avevano mai considerato, ma anche come nel campo della biomedicina la possibilità di ulteriori combinazioni degli elementi hanno permesso all'IA di scoprire nuovi farmaci la cui sperimentazione li ha accreditati come efficaci ricorrendo a processi molecolari non messi in campo dai ricercatori. Così come nelle applicazioni in campo militare e non solo nella produzione di jet pilotati/guidati da IA con capacità di combattimento superiori a quelle dei piloti meglio addestrati, ma anche nella valutazione delle informazioni utili per gestire al meglio la fase di condotta quanto di valutare modi, mezzi e opportunità dell'avversario.

Insomma, per Kissinger e gli altri autori del volume, tutto questo dimostra come l'Intelligenza Artificiale rappresenterà un fattore di cambiamento determinante per gli assetti sociali,

problemi teorici inerenti alle relazioni internazionali, che rientrano nei due problemi di causalità e grande teoria, e su come la teoria della complessità può fornire possibili soluzioni ai dilemmi che questi problemi creano.

economici e, ovviamente, politici. Con un'avvertenza però: credere che l'Intelligenza Artificiale possa risolvere ogni problema strategico in futuro potrebbe risolversi nell'essere solo un'ottimistica speranza. Tuttavia, è anche vero che ogni sistema autonomo di elaborazione di dati, ogni algoritmo realizzato da una serie di combinazioni di informazioni dotate di una logica predefinita – per permettere di giungere alla massima possibilità esprimibile di apprendimento della macchina, alla massima capacità di elaborazione di dati – consente di rispondere con valutazioni e punti di vista superiori alle stesse possibilità umane. Questo, riferito alle diverse minacce multidimensionali che riguardano la competizione e gestione dei mercati, le economie delle singole nazioni o i processi politici di governo sino a prevedere e gestire attacchi informatici a guida IA. In altre parole, per Kissinger, Schmidt e Huttenlocher,

L'Intelligenza Artificiale trascende già la percezione umana, in un certo senso attraverso la composizione cronologica o il viaggio nel tempo; alimentata da algoritmi e potenza computazionale, analizza e impara attraverso processi per completare i quali la mente umana avrebbe bisogno di decenni se non di secoli (Kissinger, Schmidt, Huttenlocher, 2023: 184 e ss.).

Insomma, l'uso dell'Intelligenza Artificiale farà sì che saranno raggiunti risultati a diversi livelli di complessità e di soluzione nei diversi settori dell'esperienza dell'esistenza umana. Ma, ciò considerato, gli autori ritengono che ci si dovrà dotare di una etica condivisa dell'Intelligenza Artificiale, nella progettazione e nell'uso.

5. Plasmare un futuro “intelligente”

Il futuro prossimo declinato in un presente tecnologico si manifesterà attraverso un modello tecnopolare a guida IA che riorganizzerà non solo i termini di relazione, ma anche i paradigmi tecnologici dei prossimi anni inaugurando una nuova era definibile come post-digitale per un verso e neo-neo-digitale se si considera il digitale non più un elemento distintivo, ma la *conditio sine qua non* per raccogliere ulteriori sfide che si trasferiranno nella competizione in materia di IA. Ciò significa che plasmare il futuro, stabilendo nuove norme su principi comuni circa l'uso di sistemi IA, rappresenterà una condizione fondamentale per non affidarsi a un'Intelligenza Artificiale che non tenga conto degli effetti e che senza un'etica potrebbe trascinare l'umanità intera verso un destino non felice in termini di garanzie di sopravvivenza della stessa umanità (Romeo, 2024). Per Yuval Noah Harari, autore di *Nexus: A Brief History of Information Networks from the Stone Age to AI*, la minaccia è meno fantastica e più insidiosa di quanto si possa ritenere a conoscenze date (Harari, 2024). Per Harari,

Per manipolare gli esseri umani, non è necessario collegare fisicamente il cervello ai computer (Harari, 2024: 514-515).

In fondo,

Per migliaia di anni profeti, poeti e politici hanno usato il linguaggio per manipolare e rimodellare la società. Ora i computer stanno imparando a farlo. E non avranno bisogno di inviare robot killer a spararci. Potrebbero manipolare gli esseri umani per premere il grilletto (Harari, 2024: ibidem).

Per l'autore di *Nexus*, le reti di informazione sono da considerarsi quali strutture di base su cui si fondano le nostre

società. Esse dimostrano che se «il potere deriva sempre dalla cooperazione tra un gran numero di esseri umani», il «collante» che tiene insieme queste reti di cooperazione non può essere che l'informazione, che molti filosofi e biologi vedono come «l'elemento fondamentale della realtà» e il carburante per le forme di IA (Harari, 2024, ibidem). Per Harari, le nuove tecnologie e, quindi, la stessa IA determina una nuova possibilità di gestire informazioni se non andare anche oltre secondo capacità di elaborazione e di qualità senzienti che la renderebbero capace di mutuare il comportamento umano con l'ambizione di poterlo guidare. A differenza delle tecnologie precedenti come libri e radio, per Harari,

L'Intelligenza Artificiale è il primo strumento in grado di prendere decisioni e generare idee da solo (Harari 2024, ibidem).

Harari spiega e molto chiaramente quanto gli algoritmi visti come agenti autonomi possano far diventare gli esseri umani in strumenti che l'IA può manipolare con una forza ancora più terribile. A meno che si agisca immediatamente, questa «intelligenza aliena», come preferisce chiamarla Harari, potrebbe innescare processi apocalittici inimmaginabili, fino alla distruzione della civiltà umana. Si tratta di una visione di certo non ottimistica dell'IA, il che giustificherebbe una certa precauzione considerato che anche l'industria dell'Intelligenza Artificiale ha iniziato a esprimere preoccupazioni nel senso. Tuttavia, pur di fronte a una visione poco ottimistica e costretti a fare i conti con una corsa alla conquista della supremazia tecnologica con la quale definire una sovranità, un'indipendenza necessaria per garantirsi la sopravvivenza politica ed economica, l'idea che si possa costruire un modello di relazioni cooperative sottende la volontà non solo di regolamentare l'uso delle nuove intelligenze e il ruolo loro affidabile, quanto il dividerne le opportunità o evitare che siano strumento di politiche egemoniche che determinerebbero un

sistema diviso tra potenze tecnologicamente dominanti e attori-Stato/consumatori/utenti egemonizzati dalla tecnologia delle prime. In altri termini, l'unica possibilità di costruire un sistema cooperativo è data dalla volontà di regolamentare gli accessi e l'uso dell'Intelligenza Artificiale come già provato dall'Unione europea che di fronte alle «nuove Intelligenze» ha cercato nel vertice di Londra del novembre 2023, AI-Safety Summit, di provare a superare ogni definizione mettendo sul tavolo delle discussioni un altro problema: l'impatto e il ruolo che l'Intelligenza Artificiale, nelle sue diverse e complesse declinazioni, assumerà nei prossimi anni. Una preoccupazione non di poco conto, al di là della sincerità dei 28 paesi firmatari più l'Unione europea, contenuti nella «Dichiarazione di Bletchley Park» (AI Safety Summit, 2023). Ciò non dovrà solo limitare il rischio di essere subordinati a politiche di colonialismo digitale e, quindi, tecnologico, ma richiederà anche di porre in essere norme e favorire comportamenti tali da evitare che si formino derive unilaterali in totale disaccordo con la politica a sostegno di relazioni cooperative e multilaterali come indicato proprio al Summit di Londra, dove il motivo centrale è stato rappresentato dalla volontà di mettere in campo tutte quelle iniziative utili a favorire una distribuzione equa di potere e di responsabilità in materia di governance dell'IA; cioè, una comune volontà di agire (*Pledge action*) (Romeo, 2024: 120). Questo, perché l'epoca che sarà si distinguerà dal passato non solo per la quasi totale dipendenza da sistemi tecnologici, ciò lo è sempre stato in fondo, ma per la velocità e il livello di sostituibilità che proprio tali sistemi hanno assunto rispetto alle sole capacità dell'uomo, implementate domani dall'IA quale moltiplicatore di capacità e di novità. I modelli competitivi che segneranno, quindi, il nostro presente, e che inevitabilmente caratterizzeranno il nostro futuro, non potranno che vedere aumentare tutte quelle attività di acquisizione di informazioni strategiche che verranno processate

dall'Intelligenza Artificiale nel quotidiano digitale. Una condizione che non può destare sorpresa se si osserva con obiettività l'evolversi delle conquiste tecnologiche nel mondo e la corsa a superare in qualità il limite raggiunto, facendo sì che l'ordine geopolitico – di fatto definito da una nuova instabilità per differenza di potenza – troverà nel controllo della competizione tecnologica nelle sue diverse forme, da quella cyber, all'Intelligenza Artificiale al Quantum Computing, la soluzione e/o la causa dei conflitti e delle crisi future.

6. Transizioni tecnologiche e nuove sovranità

Gli aspetti di una transizione verso nuovi paradigmi geopolitici, ma anche sociali ed economici, che superano la stessa globalizzazione come processo ancorandola ad una tappa necessaria dell'evoluzione delle prossimità verso una visione condivisa, trovano una loro efficace descrizione nei diversi ambiti di manifestazione delle relazioni umane (Banerjee, Erçetin 2019). Se, citando Aristotele, una minima deviazione iniziale dalla verità si moltiplica poi mille volte, l'affermare anche che l'equivalente nel mondo di oggi potrebbe essere che «le ali di una farfalla in Argentina possono causare un tornado in Messico» permette di descrivere con esattezza il livello di interdipendenza e di penetrazione dell'intimo individuale che la rivoluzione informazionale ha già realizzato nel nostro presente. In altre parole, è evidente che ogni cambiamento strutturale degli assetti sociali, determinato dall'impatto delle tecnologie, ridefinisce ogni scontata visione del passato attribuendo valore a un futuro presente. In questa prospettiva, prevedibilità o, meglio, la predittività come capacità attraverso la quale conquistare posizioni di vantaggio diventa una qualità necessaria per affermare una

supremacy fondata proprio sul differenziale tecnologico conquistato ricorrendo nei prossimi anni a processi di IA (Taylor, 2024). In questo senso, il nuovo «sistema-mondo», vedrebbe affermarsi una nuova versione della sovranità questa volta letta in termini di indipendenza tecnologica così come si ridetermineranno anche nuovi assunti nazionalistici dettati dalla corsa alla supremazia tecnologica per alcuni già rideterminati come tecno-nazionalismi. E, questo, considerato che le tecnologie utili non solo a sviluppare e produrre sistemi di IA ma anche ad implementarli sono appannaggio, per oltre la metà delle infrastrutture necessarie, di Cina e Stati Uniti. Una realtà non di poco conto se si considera non solo la possibilità che si affermi un nuovo ordine gerarchico internazionale – dove il metro di misura sarà la disponibilità di tecnologia IA e l'accesso alla stessa –, quanto il definirsi di nuove asimmetrie che di certo non favoriranno la transizione verso un modello relazionale più competitivo. I termini di eguaglianza, posti alla base del riconoscimento della sovranità come condizione di pari dignità giuridica, politica ed economica tra gli Stati e di autodeterminazione rischia di essere sovvertito da spinte tecnocratiche che sottendono la creazione di nuove dipendenze che determineranno non solo nuove morfologie – ovvero nuovi assetti organizzativi delle stesse comunità non necessariamente nazionali – ma disegneranno un'altrettanto diversa geografia dove saranno le disuguaglianze tecnologiche a definire i nuovi confini tra possibilità, esclusione e/o dipendenza (LaForge 2024). Infatti, il vero rischio per i paesi a minor capacità di accesso alle nuove tecnologie, siano esse mutate nello spazio digitale che rappresentate dalle applicazioni della IA, è quello di finire nella trappola di un colonialismo digitale che tende a rendere subalterna ogni possibile concorrenza nel settore. Nicolas Mialhe, senza cadere nella retorica, nel suo *The Geopolitics of Artificial Intelligence: the Return of Empires?* del 2018 si chiedeva se e in che

misura il rischio di un confronto prossimo, anche in termini di scontro militare, potesse essere alla base di una nuova guerra mondiale per la supremazia tecnologica considerato che la conquista delle migliori possibilità di IA rappresenterà il paradigma delle relazioni internazionali, politiche ed economiche di un domani non così lontano (Miaillhe, 2018). Anzi in questa prospettiva, Miaillhe è andato ancora più avanti delineando un «sistema-mondo» che vede aggiungersi agli Stati tradizionali gli attori non-Stato dotati di un alto potere di influenza come le Multinational Corporations (MNCs) le quali non solo scriverebbero nuove pagine di colonizzazione digitale e tecnologica in materia di IA, ma attribuisce loro modalità adattive alle stesse esperienze storiche imperiali. Tutto questo, considerando che le MNCs tenderanno non solo a controllare i mercati delle materie prime per garantirsi capacità di produzione, ma per massimizzare propri profitti e creare dipendenze non rinunceranno a concentrare la propria influenza assumendo ruoli e determinando comportamenti più assertivi dovuti alle loro decisioni più di quanto richiesto allo Stato tradizionale conquistandone la supremazia tecnologica. Miaillhe si chiede, infatti, se ancora oggi e nel domani prossimo si può ancora credere di poter utilizzare gli stessi termini che caratterizzano una realtà imperiale di fronte al presentarsi di nuovi imperi digitali considerandoli come una sorta di aggregato di multinazionali sostenute e dirette da Stati o cartelli, con capacità finanziarie importanti e tali da garantire il controllo delle attività tecnoscientifiche di ricerca e di applicazione attraverso le quali produrre economie di scala. In altre parole, se nel passato un impero si caratterizzava per l'esercizio del potere su un vasto territorio, per una disuguaglianza relativa tra il potere centrale e le "regioni" amministrate – spesso associate con la volontà di espandersi – e per l'attuazione di un progetto politico attraverso varie forme di influenza economica, istituzionale e ideologica, subordinando la

propria sopravvivenza nell'esprimere la migliore governance tra centro e periferia – oggi e domani non sarà più così (O'Neill, 2016). Per Mialhe, insomma, l'evoluzione delle formule di governance egemoniche, che vedono nell'idea imperiale la massima espressione del potere, saranno condizionate dai termini tecnologici attraverso i quali si riorganizzeranno le relazioni umane. Ciò significa, ad esempio, che se la rivoluzione digitale ha prodotto un decentramento economico con l'avvento dell'IA si rafforzerebbe, al contrario, una tendenza a investire verso la centralizzazione con il risultato di concentrare ancora di più il potere economico, militare, politico e ...tecnologico nelle mani di pochi attori non necessariamente statuali. Tutto questo, ovviamente, sfruttando al meglio quelle economie di scala che caratterizzano gli investimenti in tecnologia. Ciò significa, ancora, che si presenterebbe sull'orizzonte degli eventi una nuova polarizzazione dovuta a una divisione del mondo tra attori tecnologicamente evoluti e attori lasciati ai margini delle possibilità di accesso. In altre parole, una nuova divisione del mondo in Paesi e comunità aggregate tecnologicamente "ricche" e Paesi e/o comunità aggregate tecnologicamente "povere". Una realtà misurabile già da oggi con Stati Uniti e Cina più altri attori pubblico-privati che assumono una veste di potenze tecnologiche globali, mentre l'Unione europea segnerebbe il passo accontentandosi di ritagliarsi, per differenza, almeno una posizione di "non allineamento". Per Mialhe, inoltre, il colonialismo digitale – e la riflessione si riferisce anche all'Intelligenza Artificiale e alla sua tecnologia – si risolverebbe in un controllo dei software utilizzati dalla Corporation nell'uso di *Server as a Service* (SaaS), in un controllo dell'hardware e in un controllo del network. Il risultato sarebbe violare il principio della *net-neutrality*. In questo caso attori come le MNCs non solo continueranno a occuparsi di tecnologia digitale, ma sconfineranno e domineranno l'ecosistema tecnologico tendendo a esercitare il proprio dominio con formule

evolute di IA. In questo modo, si assicureranno profitti ricavati dalla proprietà intellettuale e dall'accesso alle infrastrutture e, potenzialmente, si garantiranno grandi capacità di sorveglianza esercitando un controllo sul flusso di informazioni, ad esempio, nella divulgazione di notizie o di servizi in streaming o sulle attività sociali, come i social network e lo scambio culturale (Romeo, 2024: 118-120).

7. *Egemonie di ritorno*

Insomma, la disponibilità di capacità tecnologiche diventa, quindi, un moltiplicatore di capacità politiche ed economiche riorganizzando in termini di superiorità di azione i fattori di potenza, se non le condizioni attraverso le quali si potrà riscrivere un nuovo modello di equilibrio se, soprattutto nel campo delle nuove intelligenze, si garantiranno pari opportunità/possibilità di accesso. D'altra parte, a condizioni geopolitiche date, la sopravvivenza di una potenza che ritiene di essere dominus di un sistema apparentemente unipolare, e le spinte ad affermare una nuova prospettiva multipolare – Cina, Russia, India e Brics nel loro complesso – farà sì che le teorie egemoniche non potranno che essere riscritte secondo formule diverse dal passato soprattutto se il *quantum* di progresso tecnologico, nel campo del digitale e dell'Intelligenza Artificiale, determinerà maggiori possibilità di scegliere azioni e assumere posizioni più vantaggiose, decidendo in tempi più rapidi rispetto ai competitor di riferimento. Infatti, a tal proposito, non sembra vi siano dubbi, riprendendo le riflessioni di Robert Gilpin, e della sua *Hegemonic Stability Theory*, nella loro generalità, che i cambiamenti di sistema avvengono per soddisfare nuovi interessi, fondati su di una ridistribuzione del potere relativo. Se così è, e lo è, si potrebbe anche affermare che in una

ridistribuzione di potere e di potenza espressa in capacità tecnologiche in una società internazionale particolarmente liquida, anche la teoria del caos che dà senso a un termine che non è solo un fenomeno matematico ma anche una parola chiave che ha appena cominciato ad essere applicata alle scienze sociali, permette di comprendere come e in che misura il potenziale che porta con sé può aiutare a interpretare quello che sarà il nuovo mondo. Un mondo, dove comportamenti e, quindi, scelte e soluzioni e/o posizioni saranno determinate, nella realtà di una società dominata dal *Quantum Computing*, dal grado di *entanglement* possibile. Cioè, di disporre di qualità tecnologiche tali da esprimere la migliore e più performante correlazione sia in senso positivo, quindi costruire relazioni cooperative, o in senso negativo, affermando una egemonia, quindi realizzando relazioni competitive se non gerarchiche (Gilpin, 1989). Un aspetto non di secondo piano nelle preoccupazioni della politica estera degli stessi Stati Uniti, ad esempio, rivolta ad assicurarsi la leadership nel campo non solo dell'IA ma anche della sfida sui computer quantistici. È allora evidente che la spinta in avanti a riorganizzare anche le proprie società affinché sostengano uno simile sforzo fa sì che, ancora una volta, si tenderà a utilizzare il proprio mercato interno per condizionare le politiche commerciali degli altri Paesi magari nel tentativo di correggere questa situazione e impedire (o disincentivare) che qualche Paese possa cadere in tentazione. Pertanto, superiorità economica o militare, in un quadro di nuova egemonia, il risultato potrebbe essere quello che vedrà il Paese egemone garantire la stabilità politica del sistema internazionale conquistando, comunque, una propria supremazia tecnologica. Il passaggio successivo non potrà che essere rappresentato dalla necessità di intervenire per consolidare tale supremazia, ad esempio, facendo diminuire gli incentivi strategici ai paesi secondari per mantenere la propria autonomia e indipendenza

economica, che sarebbe altrimenti dettata da considerazioni militari. Questo è il rischio per l'Europa nei confronti di una correzione della direzione della politica internazionale degli Stati Uniti o, se si cede al fascino della competizione portata sullo stesso campo della IA e delle capacità cyber, da parte della Cina. D'altra parte, come ricordava Gilpin, ogni sistema politico – sia domestico che internazionale – è definito per difendere e promuovere gli interessi di chi lo ha creato e la stabilità del sistema internazionale sarà garantita fin quando nessun *player* avrà un interesse/incentivo a volerlo cambiare (Gilpin, 1989). Tuttavia, egemonia o meno, è anche vero che il vero paradosso di un sistema egemonico è rappresentato dal fatto che le politiche che il Paese egemone promuove (liberi commerci e stabilità politica) finiranno per destabilizzare il sistema stesso per cui gli Stati Uniti, ma anche la Cina, proveranno a evitare che il distacco tra il Paese egemone, quali essi si percepiscono, e gli altri possa ridursi per via della convergenza condizionale; cioè, per la tendenza dei Paesi meno sviluppati a crescere a tassi superiori a quelli dei paesi sviluppati anche in termini di produzione/disponibilità/accessibilità all'IA. In questi termini, si possono ancora una volta richiamare e adattare alle condizioni date, le riflessioni di Gilpin, dal momento che se l'egemonia è una condizione nella quale un Paese è superiore agli altri in termini di risorse disponibili economiche e, quindi, anche militari, di fatto a esse si aggiunge il patrimonio di conoscenze tecnologiche e la capacità di applicarle nelle relazioni internazionali come nel quotidiano vissuto delle rispettive comunità tecnologicamente più evolute. Queste ultime, risultanti di un processo iperglobalizzante che supera ogni confine fisico attraverso la condivisione di dispositivi e relative applicazioni. D'altronde, come si potrebbe confutare, anche in questa nuova dimensione, il fatto che un Paese egemone possa costringere altri omologhi ad aprire i loro commerci, usando come “carota” l'accesso al suddetto

mercato, quindi quello tecnologico e delle relative materie prime, e minacciandone la chiusura nel caso gli altri paesi dovessero abbandonare le regole del mercato imposte dalla potenza egemone? (Keohane, 2005; Mendes 2018). Un assunto che se riletto oggi, al netto della crisi della potenza americana e nella fase più interessante di proposta di un modello multilaterale sostenuto dai membri dei Brics, sembra quasi essere un monito soprattutto se – a elezioni presidenziali conclusesi con la vittoria di Trump nel novembre 2024 – dimostra quanto e in che termini l’erosione stessa dell’egemonia a stelle e strisce aprirebbe necessariamente a un coordinamento multipolare. Questo, dal momento che se è vero che uno Stato è abbastanza potente da mantenere le regole essenziali che governano le relazioni interstatali ed è disposto a farlo, la riserva di Trump a “volarlo” fare in prospettiva risponde chiaramente al quesito precedente (Keohane, Nye, 1977; Keohane 1980). Infatti, ancora una volta, è la storia che ci ricorda ciò che spesso si dimentica. E cioè, che la presenza di una potenza egemone non sempre è indicatore di stabilità e garanzia di cooperazione internazionale. L’esempio del Regno Unito è sin troppo evidente allorché nel diciannovesimo secolo, pur disponendo di una forza militare soprattutto sui mari, non riusciva comunque a far rispettare sempre le sue politiche commerciali liberali. D’altronde, lo stesso William C. Olson, indicava molto bene quanto la stessa HST non escludesse *a priori* la possibilità di cooperazione tra pochi Stati che possono «fornirsi di beni collettivi senza fare affidamento su incentivi positivi a parte il bene stesso» (Olson, 1993). Alla fine, il risultato è che anche un Charles Kindleberger, uno dei principali sostenitori della teoria della stabilità egemonica, manca di coerenza nelle sue argomentazioni parlando di egemonia da un lato e di leadership condivisa dall’altro poiché egemonia e condivisione possono essere se non la negazione di se stesse, di certo due ossimori di sistema la cui soluzione era

già stata ben affrontata da Robert Keohane nello spiegare che una cooperazione non egemonica può realizzarsi solo attraverso l'azione collettiva e i regimi internazionali (Kindleberger, 1989: IX; Keohane 1980). E, questo, di fatto e ancora oggi rappresenta il limite della corsa a una supremazia tecnologica se ad un tempo non si disporrà di quel *know-how* necessario a innovare nei processi e se non si potrà fare affidamento alle materie prime necessarie per confezionare quel quanto di tecnologia necessario per ritagliarsi importanti spazi di autonomia e, soprattutto, di indipendenza.

8. *Cambiamenti di paradigmi*

A condizioni date, si potrebbe dire che la stessa teoria della stabilità egemonica è più da interpretarsi come una sorta di spiegazione della cooperazione di successo in un sistema internazionale fondato su una cooperazione egemonica da parte di più attori che potrebbero esprimere, a loro volta, una propria *supremacy* ma che, in realtà, comprendono che un approccio coercitivo – ovvero, una imposizione gerarchica di un ordine egemonico – sarebbe estremamente costoso nel mantenerlo in vita per anni, sia in termini di impegno politico che, soprattutto, economico. Tuttavia, è anche vero che un declino di un modello egemonico potrebbe favorire un sistema più cooperativo caratterizzato da una distribuzione del potere e delle opportunità tali da giustificare, ad esempio, il passaggio da un sistema unipolare ad un sistema multipolare e, questo, trasferito sul piano della nuova competizione tecnologica fondata sulla produzione di nuove tecnologie e di nuovi processi di governance, rappresenta la sfida di domani (Snidal 1985). Ecco, allora, che proprio nella consapevolezza di un'egemonia in declino in epoca contemporanea, il coordinamento multipolare è rilevante e particolarmente

necessario così come il riflettere su come, e su quali basi soprattutto, poter costruire modelli post-egemonici dove proprio il patrimonio di conoscenze tecnologiche farà la differenza. Così il risultato sarà, nel passaggio da un'era classica a una nuova modernità, nel come e in che modo l'umanità nelle sue diverse forme di espressione riuscirà a superare problemi sempre più complessi, avvalendosi di nuovi supporti ai processi razionali di analisi e deduzione forniti da nuove intelligenze (Floridi 2021, 2022). Non è un caso che in tempi diversi, sull'interazione tra poteri e reti Niall Ferguson in *The Square and the Tower: Networks and Power, from the Freemasons to Facebook*, dimostrava come in un arco abbastanza ampio la storia dell'umanità fosse e sia la storia di modelli di vita e di governance, politica ed economica, e di modelli culturali e tecnologici influenzati dalla costruzione di reti, dai mutamenti gerarchici nell'ordine internazionale per gli Stati, economico per gli attori finanziari e sociale per gli individui (Ferguson, 2019). Reti, dove la tecnologia ha assunto il ruolo di guida nel capovolgere modelli o nell'offrire alternative anche percettive sino a giungere alla rivoluzione digitale che ha aperto nuove frontiere sociali, creando quel cittadino digitalmente transumante definito «netizen» la cui illusione è quella di considerare democratico il modello relazionale offerto dal web⁶.

Ferguson sottolinea quanto il fermarsi a considerare acriticamente positivo il modello gerarchico offerto dalle piattaforme di rete apra le porte ad ogni eccesso e pericolo, dalla violenza terroristica alla manipolazione delle coscienze, per cui solo l'ordine gerarchico, alla fine determinerebbe le migliori condizioni per garantire un equilibrio geopolitico considerato che, tenendo conto delle vicende storiche, l'affidarsi alle reti per il governo del mondo rappresenterebbe una “ricetta perfetta per l'anarchia” (Ferguson, 2017).

⁶ Il termine «netizen» è ben espresso in Ferguson (2017).

Qualche anno dopo, Ray Dalio nella parte finale del suo *Principles for Dealing with the Changing World Order: Why Nations Succeed and Fail*, offriva alcune indicazioni per affrontare le sfide del futuro prossimo cercando di descrivere ciò che andrebbe inteso come futuro e, in particolare, il rapporto del futuro con l'esistenza umana. Ovvero, quel legame tra la conoscenza dell'esperienza e ciò che sarà dopo il presente vissuto come un futuro che sposta la sua linea volta per volta. Un aspetto importante se si considera in prospettiva quanto l'IA contribuirà a mutare anche la gestione del sapere, modificando ciò che sino a ieri era considerato come intrinsecamente conoscibile.

Il perfezionamento progressivo delle tecnologie, il sapere come i cambiamenti avvenuti in ogni epoca e determinati dalle diverse rivoluzioni non solo politiche o religiose o culturali ma anche, se non, soprattutto economiche non lasciano dubbi sul fatto che quanto avvenuto possa ripetersi con modalità nuove ma attraverso processi simili e, quindi, già vissuti. Per Dalio, insomma, per affrontare il futuro bisogna: «1. riconoscere e adattarsi agli eventi, anche se non li si può prevedere; 2. collocare le probabilità degli eventi futuri; 3. saperne abbastanza sulle probabilità future per riuscire a tutelarsi da eventi inaspettati, anche se non si riesca a farlo perfettamente» (Dalio, 2022: 451 e ss.).

9. Relazioni quantistiche

La sfida, si risolverà nella capacità di attribuire un significato a processi tecnologicamente evolutivi che determineranno decisivi cambiamenti non solo nel quotidiano, ma negli stessi rapporti di forza tra gli Stati e i nuovi attori tecnologicamente capaci di influenzarne le scelte. Ciò significa tener conto che la prospettiva di sicurezza e difesa aprirà nuovi

scenari di competizione e garanzia anche nell'uso dell'Intelligenza Artificiale vista come un'opportunità per favorire politiche di crescita, ma anche foriera di non pochi rischi nella corsa ormai globale verso una supremazia tecnologica. Il pericolo che si possano affermare interazioni senza scrupoli come il favorire il consolidarsi di coscienze algoritmiche pongono l'obiettivo di come e di quanto potranno essere affidabili strategie di sicurezza intelligenti di fronte a minacce complesse e multidimensionali che si affacciano sul futuro delle relazioni politiche ed economiche degli Stati. Questo perché i termini della competizione tecnologica non si esauriranno solo nella complementarità tra mondo cyber e IA, quanto dal come e con quali risultati ed effetti, la sfida si sposterà sul piano della tecnologia quantistica. In quest'ottica, lo stesso *Quantum Computing* rappresenterà la nuova ulteriore spinta in avanti verso la supremazia tecnologica dettata dal riuscire a superare il determinismo (considerato come misurabilità in termini matematici dei fenomeni naturali) e sfruttare le tecnologie offerte dalla fisica quantistica. In questo «futuro presente», ci si dovrà abituare a nuove narrative e nuove frontiere delle relazioni politiche viste come espressioni di un potere organizzato questa volta sulle conoscenze tecnologiche attraverso le quali si misurerà la dipendenza e/o l'indipendenza di uno Stato o di una comunità. È evidente, insomma, che per realizzare un modello possibilmente cooperativo in un quadro tendenzialmente competitivo si tratterà di riuscire a definire un sistema relazionale dove tutte le parti devono vedere la vita non come una lotta a somma zero, ma come un luogo nel quale il multilateralismo inclusivo si presenti come l'unico gioco a somma positiva caratterizzato da una volontà cooperativa diretta a promuovere il guadagno/vantaggio reciproco. Questo perché, soprattutto in modelli sociali fortemente dipendenti dalle conoscenze tecnologiche, le relazioni all'interno e tra le società, tra

gli stessi Stati, sono complesse, non lineari e per molti versi imprevedibili.

Tra dimensione cyber e progresso verso l'Intelligenza Artificiale, la corsa alla supremazia tecnologica rideclinata in supremazia quantistica non risparmierà nuove frontiere e non si lascerà intimidire da scrupoli di carattere etico. Questo, considerato che la sfida al dominio delle superintelligenze, nel campo dell'IA come nel *Quantum-Computing* determinerà nuovi assetti e nuove modalità di governance cui dovranno corrispondere nuove norme e nuove strategie rivolte a non fare della contrattazione internazionale una pericolosa occasione per politiche di potenza indirizzate alla conquista della supremazia tecnologica dove la sovranità e l'influenza sono determinate non dal territorio fisico o dalla potenza militare, ma dal controllo su dati, server e, soprattutto, algoritmi (Srivastava, 2021). In altre parole, traducendo sul piano pratico, la possibile complementarità tra *Quantum computing* e IA di certo permetterebbe di poter offrire, all'IA, informazioni man mano superiori a quelle di partenza potendo non solo processare un maggior numero di dati, ma potendosi affidare alla correlazione (*entanglement*) di informazioni acquisite e dove ogni qubit processato è il risultato della misura di un altro qubit e così via, al di là delle distanze e delle fonti. Si realizzerebbe, in sostanza, un'architettura di elaborazione dati simile a quella umana e aggiornabile man mano che sempre più dati verranno a essere acquisiti e processati correlazionalmente. L'*entanglement*, unitamente alla sovrapposizione, rappresenta il cuore del funzionamento del computer quantistico dal momento che, molto semplicemente, senza l'*entanglement*, non si potrebbero correlare i dati con le informazioni di volta in volta ricevute. Infatti, il «Quantum entanglement» (o correlazione quantistica) può essere inteso come fenomeno e come enigma allo stesso tempo, tipico della fisica quantistica che si risolverebbe nel cosiddetto

«paradosso di Einstein, Podolsky e Rosen». Nel 1935 Albert Einstein, Boris Podolsky e Nathan Rosen scrissero a riguardo un articolo dal titolo *Can Quantum-Mechanical Description of Physical Reality be Considered Complete?* conosciuto anche come «The EPR Paper» dalle iniziali degli autori.

In questo saggio Einstein, superando un certo scetticismo della prima ora verso l'approccio quantistico, e ragionando «contro-intuitivamente» rispetto all'approccio tipico della meccanica classica e di una prospettiva deterministica – ancorché ricondotta all'interno della teoria della relatività – sosteneva che non solo la teoria quanto-meccanica del mondo era da considerarsi vera, ma che da queste osservazioni doveva anche derivare il paradossale risultato definito nella possibile relazione di particelle che possono, anche senza aver mai condiviso un sistema, e a informazioni date, correlarsi riconoscendo un comune dato informazionale per il quale il mutamento dell'una avrebbe comportato anche un pari mutamento dell'altra al di là della distanza intercorrente tra le stesse (Romeo, 2024: 120). Questo nuovo spostamento in avanti rispetto alla relatività ristretta formulata anni prima, fece sì che Einstein mutasse l'opinione per la quale, in una precedente lettera a Max Born, scriveva che «Non potrò mai credere che Dio giochi a dadi con l'Universo» (Einstein, Podolsky, Rosen, 1935). Tuttavia, la fisica quantistica determina posizioni molto diverse e il dubbio di Born sembra essere risolto dal punto di vista di Stephen Hawking per il quale, al contrario,

Noi siamo il prodotto delle fluttuazioni quantistiche presenti nell'universo primordiale. Chi fosse religioso potrebbe dire che davvero Dio gioca a dadi (Hawking, Mlodinow, 2017).

Concludendo

Se gli Stati-nazione hanno definito l'ordine globale da quando il Trattato di Westfalia li ha consacrati come elementi costitutivi della geopolitica quasi 400 anni fa, oggi, al contrario, poche grandi aziende tecnologiche come Amazon, Alibaba, Alphabet, Apple, Meta, Microsoft, Tencent, Tesla, quanto gli aggregati finanziari come Blackrock, State Street, Vanguard, sono effettivamente diventati soggetti indipendenti e sovrani nei vasti e crescenti regni digitali che hanno creato e controllano e, per questo, stabiliscono regole ed esercitano potere nello spazio metaumano proprio come fanno i governi nazionali sul territorio fisico. Ma non solo. Hanno creato anche una nuova dimensione geopolitica; cioè, uno spazio digitale nel quale esprimono tutta la loro capacità di persuasione dal momento che sempre più persone vivono il loro quotidiano all'interno di tale spazio dematerializzato, dove con non poche difficoltà gli stessi governi cercano di controllare il consenso. Le Big Tech come le realtà dell'e-commerce, hanno dimostrato di riuscire a esercitare un potere, e un'influenza, paragonabile a quella dei principali Stati giustificando anche la produzione di politiche di manipolazione per creare una coscienza collettiva direzionabile sugli obiettivi del dominus politico ed economico del momento, introducendo quel capitalismo di sorveglianza che tende ad assicurarsi non solo il controllo dei mercati e delle decisioni ad esso riconducibili, ma anche il controllo dei consumatori dirigendone le scelte (Zuboff, 2019). Un avvertimento che potrebbe lasciare sorpresi coloro che credono in un approccio più realisticamente cooperativo se non ingenuamente non competitivo.

Alla fine, è altrettanto vero che esiste un problema di incongruità tra le norme di uso delle piattaforme di rete per l'utente e per la *community* e le norme tradizionali, ponendo seri

problemi sul rispetto non solo di norme e prassi deontologiche, ma degli stessi diritti del singolo laddove i privilegi di partecipazione sono dettati dal gestore della piattaforma. In questo senso e non solo, l'Unione europea dovrebbe evitare di dipendere da piattaforme di rete collocate in un altrove possibile, che ha precise se non condizionanti caratteristiche geopolitiche che possono limitarne la libertà di azione, la sua competitività o comprimere i diritti dei propri cittadini. Il vero problema è che l'Unione non ha ad oggi proprie piattaforme di rete globali come invece accade negli Stati Uniti piuttosto che in Cina o anche in Russia e, nonostante tutto, ciò rischia di rendere solo una formalità la stessa adozione dell'AI-Act dal momento che le redini del gioco nei vari settori strategicamente importanti nella sfida economica sembrano aver preferito l'Estremo Oriente e gli Stati Uniti; questi ultimi quale, oggi, unica sintesi occidentale disponibile per affermare una certa capacità in termini di competizione con i giganti AI-Tech asiatici (European Parliament, 2023; Bertuzzi 2023). L'adozione di un primo testo normativo di disciplina dell'uso dell'IA non chiude un cerchio se non si afferma una competitività nel porsi tra gli attori che possono gestire le offerte in tale settore e che decideranno gli assetti di governance economica, quindi, anche politica dei prossimi anni. È quindi evidente che gli Stati e le leadership saranno costretti a dover agire in aree di possibile convergenza caratterizzate dall'intersezione tra molti nuovi mondi in cui gestiranno realtà fisiche e virtuali, realizzando *shift* geopolitici da una prospettiva globale ad un'altra in ragione degli interessi di potenza tecnologica che si affermerà negli anni a venire. In fondo, Stati e governi, a norme internazionali possibili o da rimettere in campo, saranno costretti a far sì che l'innovazione e gli standard tecnologici si intreccino con i processi di governance digitale se non addirittura affidati a sistemi e processi creati e guidati dall'IA.

Probabilmente la prospettiva di Isaac Asimov non era aleatoria per il quale visto che

È il cambiamento, il cambiamento continuo, il cambiamento inevitabile, che è il fattore dominante nella società odierna. Nessuna decisione sensata può essere più compiuta senza tenere conto non solo del mondo come è ora, ma di come sarà [...]. Questo, a sua volta, significa che i nostri uomini di Stato, i nostri imprenditori, i nostri uomini comuni, devono orientarsi verso un modo di pensare fantascientifico (Asimov, Holdstock, 1978: 5).

Tra il fascino di un nuovo futurismo si tratterà di evitare che si affermi un paradigma antropo-tecnocratico che ponga l'IA quale esemplificazione moltiplicativa delle possibilità umane in chiave anche metaevolutiva, che dovrebbero garantire quel desiderio di perfezionamento postumano tipico degli ultimi anni del Novecento e che caratterizzeranno gli anni del prossimo futuro⁷. Un perfezionamento che, attraverso la IA sembra potersi realizzare in quella figura dell'oltreuomo, volta a superare ogni limite fisico nel dominio del proprio essere e nella realizzazione di un sé tecnologicamente determinato attraverso repliche in entità non più solo, e semplicemente, senzienti.

Bibliografia

AI-Safety Summit (2023), *The Bletchley Declaration, Countries Attending the AI Safety Summit*, November 1-2, 2023, Policy Paper. <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-aisafety-summit-1-2-november-2023>.

⁷ Sull'«antropotecnica», vedasi Sloterdijk (2009).

Amir, D.A (2002). *Entanglement: The Greatest Mystery in Physics*. New York: Basic Books. (tr. it. *Entanglement. Il più grande mistero della fisica*, Milano, Raffaello Cortina, 2024).

Argyle, R., Kohara, T. (1985). Complexity in International Relations. In Denig, E., van der Meiden A. (eds), *A Geography of Public Relations Trends*. Dordrecht: Springer. 213-219. https://doi.org/10.1007/978-94-017-0575-2_19.

Asimov, I., Holdstock, R. (1978), *My Own View, The Encyclopedia of Science Fiction*. London: Octopus Books.

Banerjee, S., Erçetin, Ş.Ş. (2019). *Chaos and Complexity Theory in World Politics*. London-New York: Springer.

Bertuzzi, L. (2023). L'Unione europea trova la quadratura del cerchio con il primo regolamento sull'IA al mondo. *Euractiv Italia*, 11 dicembre, <https://euractiv.it/section/digitale/news/lunione-europea-trova-la-quadratura-del-cerchio-con-il-primo-regolamento-sullia-al-mondo/>.

Bremmer, I. (2021). The Technopolar Moment. How Digital Powers Will Reshape the Global Order. *Foreign Affairs*, 100(6), 112-128.

Clemens, W. (2013). *Complexity, Science and World Affairs*. New York: Suny Press.

Dalio, R. (2022). *I principi per affrontare il nuovo ordine mondiale. Dal trionfo alla caduta delle nazioni*. Milano: Hoepli (tit. or. *Principles for Dealing with the Changing World Order: Why Nations Succeed and Fail*, New York, Avid Reader Press/Simon & Schuster, 2021).

Einstein, A., Podolsky, B., Rosen, N. (1935). Can Quantum-Mechanical Description of Physical Reality be Considered Complete? *Physical Review*, 47, 777-780, consultabile in <https://cds.cern.ch/record/405662/files/PhysRev.47.777.pdf>.

Eurasia Group (2023). *Technopolarity*. Eurasia Group – Policy First, August. <https://www.eurasiagroup.net/files/upload/Technopolarity.pdf>.

European Parliament (2023). Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI, Press Release, 9 December; <https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai>.

Ferguson, N. (2017). The False Prophecy of Hyperconnection: How to Survive the Networked Age. *Foreign Affairs*, 96(5), 68-79.

Ferguson, N. (2019). *La piazza e la torre*, Milano: Mondadori. (tit. or. *The Square and the Tower: Networks and Power, from the Freemasons to Facebook*, New York, Penguin Press, 2018).

Floridi, L. (eds.) (2021). *Ethics, Governance, and Policies in Artificial Intelligence*, Cham: Springer.

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

Gilpin, R. (1989). *War & Change in World Politics*. Cambridge: Cambridge University Press (tr. it., *Guerra e mutamento nella politica internazionale*, Bologna, Il Mulino, 1989).

Harari, Y.N. (2024). *Nexus. Breve storia delle reti di informazione dall'età della pietra all'IA*, Milano: Bompiani (tit. or. *Nexus: A Brief History of Information Networks from the Stone Age to AI*, New York, Random House, 2024).

Hawking, S., Mlodinow, L. (2017). *Il grande disegno*. Milano: Mondadori.

Hemchi, M. (2022). *Complexity Theory of International Relations: An Introduction*. The Arab Center for Research and Policy Studies.

Kennedy, P. (2001). *Ascesa e declino delle grandi potenze*. Milano: Garzanti (tit. or. *The Rise and the Fall of the Great Powers*, New York, Random House, 1987).

Keohane, R.O. (1980). The Theory of Hegemonic Stability and Changes in International Economic Regimes, 1967-1977. In Holsti O. e al. (1980). *Change in the International System*. London: Routledge. 131-162.

Keohane, R.O. (2005), *After Hegemony. Cooperation And Discord In The World Political Economy*. Princeton-Oxford: Princeton University Press.

Keohane, R.O., Nye, J.S. (1977). *Power and Interdependence: World Politics in Transition*. Boston: Little, Brown.

Khanna, P. (2017a). *Connectography. Le mappe del futuro ordine mondiale*. Roma: Fazi Roma, pp. 447 – 530 (tit. or. *Connectography: Mapping the Future of Global Civilization*, New York, Random House, 2016).

Khanna, P. (2017b). *La rivincita delle Città-Stato*, Roma: Fazi (tit. or. *Technocracy in America. Rise of the Info-State*, North Charleston, SC, CreateSpace, 2017).

Kindleberger, C.P. (1989). *Economic Laws and Economic History*. Cambridge: Cambridge University Press.

Kissinger, H.A., Schmidt, E., Huttenlocher, D. (2023), *L'era dell'Intelligenza Artificiale. Il futuro dell'identità umana*, Milano: Mondadori (tit or. *The Age of AI and Our Human Future*, Boston, Little Brown & Co, 2022).

LaForge, G. (2024). The Dangers of Imposing Global North Approaches to AI Governance on the Global South, *Tech Policy Press*. September 5. <https://www.techpolicy.press/the-dangers-of-imposing-global-north-approaches-to-ai-governance-on-the-global-south/>.

Mendes, M.V.I. (2018). Is it the end of North-American hegemony? A structuralist perspective on Arrighi's systemic cycles of accumulation and the theory of hegemonic stability. *Brazilian Journal of Political Economy*, 38(3) (152), 434-449. <https://doi.org/10.1590/0101-35172018-2799>.

Mialhe, N. (2018). The Geopolitics of Artificial Intelligence: the Return of Empires? *Politique étrangère*, 3, 105-117.

Olson, W.C. (1993). *The Theory and Practice of International Relations*. Upper Saddle River: Prentice-Hall.

O'Neill, D.I. (2016). Empire. *Encyclopædia Britannica*. In: <https://www.britannica.com/topic/empire-political-science>.

Peyron, L. (2021). I diritti umani nella condizione digitale: un'opportunità di rilanci. In Larghero E., Zeppegno G. (a cura di). *Diritti umani e dignità nell'era della globalizzazione. Percorsi bioetici*. Cantalupa (Torino): Effatà.

Romeo, G. (2024). *Dalla dimensione cyber alle “nuove” intelligenze. Rischi e sfide per l'Europa*, Jean Monnet Centre d'Excellence, AI4AI. Università di Torino https://www.ledizioni.it/download/65426/?tmstv=1716807622&_gl=1*f63dbt*_ga*NTQ4Mjk3NTg5LjE3NDQ1NzYxNDE.*_up*MQ.

Sloterdijk, P. (2009). *You Must Change Your Life: On Anthropotechnics*. Cambridge: Polity Press. (tr. it. *Devi cambiare la tua vita. Sull'antropotecnica*, Milano, Raffaello Cortina, 2010).

Snidal, D. (1985). The Limits of Hegemonic Stability Theory. In *International Organization*, 39(4), 579-614.

Srivastava, S. (2023) Algorithmic Governance and the International Politics of Big Tech. *Perspectives on Politics*, 21(3): 989-1000. doi:10.1017/S1537592721003145.

Taylor, N. (2024). Artificial Intelligence and Its Potential to Transform Global Power Dynamics. *International Journal of Business Management and Visuals*, 7(1). Consultabile in: <https://ijbmvc.com/index.php/home/article/view/95>.

Van Dijck, J., de Winkel, T., & Schäfer, M. T. (2021). Deplatformization and the governance of the platform ecosystem. *New Media & Society*, 25(12), 3438-3454. <https://doi.org/10.1177/14614448211045662>.

Wallerstein, I. (1997), *L'era della transizione. Le traiettorie del sistema-mondo*, Trieste: Asterios, (tit. or. *The Age of Transition: Trajectory of the World System, 1945-2025*, London, Zed Books, 1996).

Wallerstein, I. (2004). *World-Systems Analysis: An Introduction*. Durham: Duke University Press.

Wallerstein, I. (2007), *La retorica del potere. Critica dell'universalismo europeo*. Roma: Fazi (tit. or. *European Universalism. The Retic of Power*, New York, New Press, 2006).

Zuboff, S. (2019). *The Age of Surveillance Capitalism: the Fight for a Human Future at the New Frontier of Power*, New York: PublicAffairs. (tr. it. *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, Roma, Luiss University Press, 2019).

QUADERNI DI COMUNITÀ del DiTES

N. 1/2021

L'ISTRUZIONE, IL LAVORO E LA SOCIETÀ AI TEMPI DELL'EMERGENZA
PANDEMICA GLOBALE

A cura di Stefania Capogna, Angelo Del Cimmuto, Concetta Fonzo

N. 1/2022

SFIDE E OPPORTUNITÀ PER I SISTEMI DI ISTRUZIONE, FORMAZIONE
E LAVORO NEL PERIODO EMERGENZIALE

A cura di Stefania Capogna, Donatella Cannizzo, Concetta Fonzo

N. 2/2022

IL PRESENTE E IL FUTURO DELLA SOCIETÀ DIGITALE. LUCI E OMBRE
DI UNA INEVITABILE TRANSIZIONE

A cura di Angelo Del Cimmuto, Fulvio Oscar Benussi

N. 3/2022

L'EDUCAZIONE E LA SOCIETÀ NEL FUTURO POST PANDEMIA

A cura di Speranzina Ferraro, Eugenio De Gregorio, Lavinia Cicero

N. 1/2023

ACTIVE CITIZENSHIP FOR THE DIGITAL SOCIETY. EXPERTISE, BEST
PRACTICES AND TEACHING IN THE DIGITAL ERA

Edited by Stefania Capogna, Manuela Minozzi, Danila Scarozza

N. 2/2023

TEACHING ENHANCED LEARNING FOR ENGAGING AND INCLUSIVE
LEARNING

Edited by Ida Cortoni, Veronica Lo Presti, and Eleonora Sparano

N. 3/2023

REINVENTING UNIVERSITY: THE DIGITAL CHALLENGE IN HIGHER
EDUCATION

Edited by Stefania Capogna, Ligita Šimanskienė, Erika Župerkienė



N. 1/2024

LA REPUTAZIONE NELL'ERA DIGITALE. RAPPRESENTAZIONI E
PRATICHE DEL SÉ TRA CAPITALE SOCIALE E BENE RELAZIONALE

A cura di Eleonora Sparano, Nicola Strizzolo, Martina Lippolis

N. 2/2024

ORIENTAMENTO AL FUTURO

A cura di Speranzina Ferraro, Lavinia Cicero, Diego Boerchi, Andrea
Zammitti

N. 3/2024

HUMAN FLOURISHING FOR WELLBEING IN SOCIETY, COMMUNITIES
AND ORGANIZATIONS

Edited by Eugenia Blasetti, Cecilia Costa, Maria Chiara De Angelis,
Eugenio De Gregorio, Andrea Velardi

N. 1/2025

HUMAN-CENTRIC APPROACH TO ARTIFICIAL INTELLIGENCE

Edited by Marco Filoni, Filippo Maria Giordano, Giorgio Grimaldi

Eurilink University Press, nata nel 2006, è la casa editrice dell'Università degli Studi "Link", Università italiana non statale legalmente riconosciuta, con forte vocazione internazionale.

Mission della casa editrice è quella di creare un'azione di "ponte" culturale tra il lettore e la società globale, l'università e il mondo del lavoro, i cittadini e le Istituzioni, proponendo pubblicazioni di elevato spessore, umano e culturale, con contenuti chiari e approfonditi, in grado di affrontare tutte le tematiche emergenti del nostro tempo e le prospettive dibattute a livello internazionale.

Eurilink pubblica in lingua italiana e in lingua inglese utilizzando anche lo strumento di diffusione informatica.

Le linee guida editoriali sono individuate da Comitati Scientifici e dalle Direzioni Editoriali delle diverse collane. Le pubblicazioni dedicate alla Formazione Universitaria e alla Ricerca Scientifica, sono sottoposte a sistema di peer review.

L'offerta editoriale si articola in 5 grandi aree, alle quali afferiscono 19 differenti collane, e comprende Riviste, Annali, Quaderni e Coedizioni:

A. ACCADEMIA: COLLANE CON FOCUS UNIVERSITARIO

- *Studi e dialoghi giuridici*: norma e diritto vivente. Ambito Privatistico (*Critical Studies in Private Law*) e Ambito Pubblicistico, Comunitario e Internazionale (*Critical Studies in Public Law, Community and International field*). Collana specialistica di riferimento del Corso di Laurea a ciclo unico di Giurisprudenza
- *Studi politici, economici, diplomatici e internazionali*: analisi dei processi di globalizzazione dei fenomeni politici e delle relazioni tra i Paesi; collana di riferimento dell'omonimo Corso di Laurea e di Laurea Magistrale
- *Studi di comunicazione digitale*: innovazioni nei mezzi di comunicazione e pubblicità; collana di riferimento dell'omonimo Corso di Laurea e di Laurea Magistrale
- *Studi di management, finanza e fiscalità dell'impresa*: analisi e metodologie; collana di riferimento dell'omonimo Corso di Laurea e di Laurea Magistrale
- *Campus*: collana che riflette il dibattito culturale e scientifico dell'Università e della Fondazione, oltre che le attività di formazione professionale, universitaria e post universitaria e i risultati delle attività di ricerca (*Handbook, Conference proceedings, Ricerca*)
- *Alumnia*: pubblicazioni dei laureati che hanno ottenuto la dichiarazione di "dignità di stampa" da parte delle relative Commissioni di Laurea
- *I codici*: raccolta di norme per le varie aree del diritto, italiano e internazionale

B. ATTUALITÀ INTERNA E INTERNAZIONALE

- *Tempi Moderni*: attualità politica, economica e sociale - italiana e internazionale
- *Ricostruzione*: analisi storica della strategia di ricostruzione morale e civile dell'Italia del dopoguerra per interpretare le dinamiche sociali e politiche del nostro tempo
- *Eurinstant*: temi emergenti rappresentati con dati e informazioni per una lettura veloce

- *La Critica*: analisi e interpretazione dei fenomeni culturali e sociali del nostro tempo
- *I Saggi*: approfondimenti monografici nelle diverse discipline

C. STORIA

- *Tracce*: ritratti e biografie, racconti e percorsi di vita
- *Historia*: studi storici, in grado di fornire contributi, anche inediti, alla comprensione di fatti e di relazioni che costituiscono le radici storiche della nostra civiltà

D. ARTE E CULTURA

- *Arti e Tradizioni*: manifestazioni dell'arte e rappresentazione delle tradizioni delle regioni d'Italia
- *Ambiente Urbano e Territorio*: analisi, studi e progetti
- *Benessere e Salute*: teoria, ricerca e percorsi operativi
- *Link*: pubblicazioni con specifica attenzione alla lingua e alla cultura di paesi emergenti

E. ISTITUZIONI

- *Pubblicazioni di servizio*: per Enti o Istituzioni varie, interne e internazionali

All'offerta editoriale pubblicata nelle diverse collane si aggiungono:

1. RIVISTE

- *Quaderni di Comunità - Persone, Educazione e Welfare nella società 5.0*, sullo studio e la comprensione della complessità culturale, sociale, organizzativa ed educativa del nostro tempo (*print and on line*)

2. COEDIZIONI: pubblicazioni in collaborazione con altre case editrici - non italiane - finalizzate a una più estesa divulgazione anche all'estero

3. ANNALI

- *Annali del CERSIG* (Centro di Ricerca sulle Scienze Giuridiche)
- *Annali del DISEC* (Centro di Ricerca "Dinamica dei Sistemi Economici Complessi")
- *Annali dell'Università degli Studi "Link Campus University"*



Il Centro di Ricerca **DITES** (*Digital Technologies, Education and Society*) dell'Università degli Studi "Link" si propone come spazio di ricerca e sperimentazione per attori, istituzioni e organizzazioni che hanno a cuore la centralità dell'educazione in tutte le sue forme: (educazione, formazione, stage, lavoro, accompagnamento, edutainment).

L'obiettivo generale del Centro di Ricerca **DITES** è valorizzare le tecnologie digitali in un'ottica di empowerment individuale, organizzativo e di comunità, mettendo la *persona* al centro dei processi di innovazione.

DITES si muove su due macro aree di interesse principali, ma non esclusive:

- a) tecnologie digitali, istruzione, organizzazioni e società;
- b) *empowerment* per persone, organizzazioni e comunità.

Confluiscono attorno al centro numerosi studiosi, esperti, formatori, nazionali e internazionali che ne condividono la filosofia.



L'Università degli Studi "Link", Università non Statale legalmente riconosciuta dall'Ordinamento Universitario Italiano ai sensi del DM 21 settembre 2011, pubblicato in G.U. Serie Generale n. 268 del 17 novembre 2011. Già sede a Roma dell'antica Università di Malta, è stata la prima Università straniera autorizzata a operare in Italia (D.M. 27.11.1999).

L'Università degli Studi "Link" nasce come una "unità scientifica" internazionale, multidisciplinare, interdisciplinare, focalizzata sui cambiamenti e sulle sfide che, a livello globale, investono i diversi ambiti della conoscenza e della ricerca.

Obiettivo generale è quello di rendere la nuova classe dirigente capace di operare in termini di "problem solving" nelle aree del sapere economico, giuridico e politico, con grande sensibilità e capacità creativa e comunicativa.

L'approccio comparatistico dello studio dell'economia, del diritto e dell'amministrazione della politica, consente di creare un laureato con un profilo integrale sia sul piano della connessione fra culture e lingue diverse sia su quello dell'interdipendenza fra l'area pubblica e quella privata, per preparare la nuova classe dirigente a operare nei diversi sistemi e Paesi, garantendo ai giovani possibilità vincenti nella mobilità del mercato globale del lavoro.

L'Università degli Studi "Link" ha una forte vocazione internazionale che realizza attraverso gli accordi stretti con numerose università estere al fine di favorire la mobilità di docenti e studenti.

L'offerta formativa dell'Università degli Studi "Link" comprende:

Corsi Triennali

DAMS - Produzione audiovisiva e teatrale (L-3); Gestione aziendale - Global management (L-18); Tecnologie innovative per la comunicazione digitale - Innovative technologies for digital communication (L-20); Scienze della politica e dei rapporti internazionali - Political sciences and international relations (L-36); Professioni Sanitarie (L-snt): Infermieristica, Fisioterapia, Tecniche di laboratorio biomedico, Tecniche di radiologia medica per immagini e radioterapia, Osteopatia; Scienze della difesa e della sicurezza (L/DS) (Sede di Napoli); Ingegneria Informatica (L-8) (Sede di Napoli).

Corsi Magistrali

Economia dell'innovazione e della globalizzazione sostenibile - Economics of innovation and sustainable globalization (LM-56); Tecnologie e linguaggi della comunicazione - Technologies and languages of communications (LM-59); Studi strategici e politiche della sicurezza - Strategic studies and security policies (LM-62); Gestione aziendale avanzata - Advanced global management (LM-77); Management del lavoro, del welfare e dei servizi sanitari - Managing of labor, welfare, and health services (LM-56).



Corsi a Ciclo Unico

Giurisprudenza (LMG-01); Medicina e Chirurgia (LM-41); Farmacia (LM-13); Odontoiatria (LM-46) (Sede di Città di Castello); Scienze della formazione primaria (LM-85 bis) (Sede di Città di Castello).

Master di Primo Livello

Brand Manager dell'Enogastronomia; Cybersecurity; Esperti per la gestione delle crisi e la cooperazione post-conflitto; MBA in Diritto e Management dello Sport; Musica e Arti dello Spettacolo; Psicologia del Cambiamento; Service Innovation & Digital Transformation; Smart Public Administration.

Mater di Secondo Livello

Anticorruzione e Appalti Pubblici. Un nuovo modello di etica pubblica: risposte ordinamentali e nuovi protagonisti; Gestione dei Beni Culturali; Intelligence and Security; Intelligence Specialist; Musica e Arti dello Spettacolo; Politiche Attive, di Direzione e Gestione delle Risorse Umane; Programmazione e Progettazione dei Finanziamenti Europei 2021-2027. Next Generation Italia; Sicurezza Ambientale: Tecnologie Innovative, Droni e Geomatica per la Tutela dell'Ambiente e del Territorio.

Finito di stampare nel
da
per i tipi di Eurilink University Press