

I quaderni di  
**Agenda  Digitale.eu**

Agendadigitale.eu è una testata scientifica e giornalistica registrata al Tribunale di Milano

### **Dati di riferimento**

Iscrizione ROC n. 16446

ISSN 2421-4167

Numero registrazione 1927, Tribunale di Milano Editore: Digital360

### **Focus e ambito**

La rivista scientifica, i Quaderni di Agendadigitale.eu, pubblica fascicoli quadrimestrali in open access.

Lo scopo è creare un luogo per accompagnare i passi dell'Italia verso la necessaria rivoluzione digitale, con approfondimenti multidisciplinari a firma di esperti delle materie afferenti all'Agenda Digitale italiana ed europea

### **Submission e norme editoriali**

Per effettuare una submission è necessario concordare prima un argomento e le misure precise contattando [info@agendadigitale.eu](mailto:info@agendadigitale.eu).

Inviare un abstract di circa 500 caratteri alla testata, presentando l'articolo.

Le misure del testo finale saranno comprese tra 6mila e 20mila caratteri, salvo accordi per misure superiori.

I riferimenti bibliografici dovranno essere preparati in conformità alle regole dell'APA style, 6a edizione (si vedano le linee guida e il tutorial).

Gli autori sono invitati a tener conto degli articoli già pubblicati nella rivista e di citarli nel loro contributo qualora siano ritenuti di interesse per il tema trattato.

## **Direzione e comitato editoriale**

### **Direttore responsabile**

**Alessandro Longo, Direttore responsabile Agendadigitale.eu**

### **Direttori scientifici**

**Paolo Ferri**, Professore Ordinario di Tecnologie della formazione,

Università degli Studi Milano-Bicocca

**Mario Morcellini**, Professore ordinario emerito in Sociologia della

Comunicazione e dei Media digitali, Sapienza Università di Roma

### **Comitato scientifico Area XIV, Scienze Politiche e Sociali**

#### **Membri del Comitato scientifico**

**Davide Bennato**, Università di Catania

**Giovanni Betta**, Consiglio Universitario Nazionale, già Rettore

Unicassino

**Bruno Botta**, Rettore Unitelma Sapienza

**Giovanni Boccia Artieri**, Università di Urbino

**Giovanni Cannata**, Rettore Universitas Mercatorum

**Marco Centorrino**, Università di Messina

**Ida Cortoni**, Sapienza Università di Roma

**Stefano Cristante**, Università del Salento

**Marino D'amore**, Unicusano

**Lelio de Michelis**, Università dell'Insubria

**Marco del Mastro**, Unicusano e Agcom

**Stefania Fragapane**, Universitas Mercatorum di Roma

**Gino Frezza**, Università di Salerno

**Gevisa La Rocca**, Università Kore di Enna

**Nicola La Sala**, Agcom; industria4.0, editoria, cittadinanza digitale

**Andrea Lombardinilo**, Università di Roma Tre

**Mariano Longo**, Università del Salento

**Michela Luzi**, Unicusano

**Michele Mirabella**, conduttore televisivo e professore a contratto di divulgazione scientifica

**Mario Morcellini**, Sapienza Università di Roma

**Simone Mulargia**, Lumsa

**Barbara Maussier**, Seconda Università di Roma

**Antonella Napoli**, Università di Salerno

**Marco Navarra**, Università di Salerno

**Sebastiano Nucera**, Università di Messina

**Achille Pierre Paliotta**, INAPP; disinformazione, IA, sociologia economica

**Francesco Pira**, Università di Messina

**Antonio Preiti**, sondaggista e organizzatore culturale

**Antonio Rafele**, Sapienza Università di Roma

**Christian Ruggiero**, Sapienza Università di Roma

**Nicola Strizzolo**, Università di Teramo

**Franco Torcellan**, Associazione RED – Laboratorio di Ricerca

“Formare Trasformare Innovare”

**Elena Valentini**, Sapienza Università di Roma

### **Comitato editoriale**

**Giovanni Boccia Artieri, Marco Centorrino, Antonio Opronolla,**

**Mario Pireddu, Luca Toschi**

**Comitato di referaggio**

**Coordinatore: Luca Gastaldi, Polimi**

**Mauro Andreolini**, sicurezza informatica, Unimore

**Luca Baccaro**, concorrenza, diritto comunicazioni elettroniche e dei media; studio legale Lipani Categorical & Partner

**Raffaello Balocco**, IT e innovazione, Politecnico di Milano

**Francesco Capparelli**, privacy, cyber security, ecommerce, data management, identità digitale; studio legale ICT Legal Consulting

**Antonio Chella**, ingegneria informatica, intelligenza artificiale, Università di Palermo

**Marco Centorrino**, Università di Messina – processi culturali e comunicativi, nuove tecnologie

**Ida Cortoni**, media education e digital literacy; Dipartimento di Comunicazione e Ricerca Sociale, Sapienza Università di Roma

**Giuseppe D'Acquisto**, Autorità garante privacy, sicurezza e privacy

**Mario dal Co**, Economista e manager, già direttore dell'Agenzia per l'innovazione

**Lelio Demichelis**, Università Insubria, sociologia, economia

**Francesco Di Giorgi**, diritto dell'informazione e della comunicazione, tutela dei consumatori, diritto delle comunicazioni elettroniche; Agcom

**Leonella Di Mauro**, data management, e-commerce, tutela del consumatore, diritto delle comunicazioni elettroniche; Agcom

**Luisa Franchina**, cyber security, Hermes Bay

**Luca Gastaldi**: eGov, sanità, telecomunicazioni, procurement pubblico, design thinking, Smart Working, Politecnico di Milano

**Maurizio Gentile**, professore associato, Università di Roma LUMSA, didattica e pedagogia

**Antonio Ghezzi**: strategia, business model, startups, mobile, Politecnico di Milano

**Gevisa La Rocca**, Università Kore di Enna, piattaforme digitali, communication research, analisi qualitativa dei dati

**Nicola La Sala**, registro degli operatori della comunicazione, fattura elettronica, industria4.0, editoria, cittadinanza digitale; Agcom

**Emanuele Lettieri**, sanità Politecnico di Milano

**Maria Beatrice Ligorio**, psicologia, università di Bari

**Marika Macchi**, economia, Unifi

**Riccardo Mangiaracina:** fatturazione elettronica, eCommerce, logistica e trasporti, export, Politecnico di Milano

**Mirco Marchetti,** Sicurezza informatica, unimore

**Chiara Marzocchi,** economia, Università di Manchester

**Cristina Masella,** Sanità, Politecnico di Milano

**Carmelina Maurizio,** Dipartimento di Filosofia e Scienze dell'Educazione Università di Torino

**Stefano Moriggi,** scienze della comunicazione, filosofia, Bicocca di Milano

**Davide Mula,** sanità digitale, cyber security, privacy; Agcom

**Simone Mulargia,** internet and social media studies; Lumsa

**Antonella Napoli,** sociologia, media e comunicazione, giornalista

**Sebastiano Nucera,** Università di Messina, Media e Tecnologie Indossabili

**Achille Pierre Paliotta,** Social cybersecurity, disinformazione, tecnologie digitali, intelligenza artificiale, sociologia economica; INAPP

**Francesco Paoletti,** docente di organizzazione aziendale e gestione delle risorse umane, Università degli Studi di Milano-Bicocca

**Norberto Patrignani,** computer ethics, filosofia, Politecnico di Torino

**Dunia Pepe,** Inapp e Università Roma Tre, cultura e formazione digitale

**Alessio Plebe,** Università di Messina, Scienze cognitive, pedagogiche, psicologiche

**Francesco Pira,** Unime, comunicazione pubblica, le dinamiche social, le fake news e i processi di disinformazione

**Franco Pizzetti,** diritto, privacy, università di Torino

**Barbara Quacquarelli,** scienze umane e formazione, università Milano Bicocca

**Antonio Rafele,** Sociologia dei processi culturali e comunicativi, Unicusano

**Filippo Renga:** turismo digitale, smart agrifood, finance and banking, mobile, Politecnico di Milano

**Angelo Rovatti,** tutela del diritto d'autore, diritti connessi, Diritto dei media; Agcom

**Christian Ruggiero,** sociologia del giornalismo e comunicazione politica; Dipartimento di Comunicazione e Ricerca Sociale, Sapienza Università di Roma

**Franco Torcellan,** Associazione RED – Laboratorio di Ricerca Educativa e Didattica “Formare Trasformare Innovare”

**Angela Tumino:** Internet of Things, logistica e trasporti, smart city, Politecnico di Milano

**Simone Vannuccini,** economia, SPRU

**Francesco Varanini,** filosofia, formazione, università di Pisa

**Guido Vetere,** Università Sapienza di Roma, intelligenza artificiale, tecnologia

## INDICE DEL FASCICOLO

Gli LLM pensano? Come funziona davvero il “cervello” digitale dell’IA - di <b>Riccardo Manzotti</b> , Ordinario di Filosofia Teoretica, IULM, Milano .....	9
Agency e tecnologie cognitive. Uno sguardo al rapporto tra IA e lavoro – di <b>Sebastiano Bavetta</b> , Dipartimento SEAS, Università degli Studi di Palermo.....	22
Conversazioni artificiali in psicologia: strumenti, illusioni e nuove domande cliniche – di <b>Raffaella I. Rumiati</b> , SISSA Trieste .....	33
Le competenze chiave per l’intelligenza artificiale generativa: verso un approccio <i>human-centered</i> al governo della AI – di <b>Giovanni Battista Dagnino</b> , Professore ordinario di Digital Strategy & Artificial Intelligence .....	39
Giornalismo e verità: chi controlla il racconto controlla il reale - di <b>Francesca Rizzuto</b> , Università degli studi di Palermo .....	45
Algoritmi editoriali. Una mappa per leggere l’automazione nelle newsroom – di <b>Marzia Antenore</b> , Dipartimento di Comunicazione e Ricerca Sociale Sapienza – Università di Roma .....	54
Bulimia social: quando l’iperconnessione diventa malattia – di <b>Marino D’Amore</b> , Docente di Sociologia generale presso Università degli Studi Niccolò Cusano.....	64
L’IA ci rende stupidi? Cosa dice davvero la scienza – di <b>Marco Giacomazzi</b> , Università di Bologna .....	69
La GenAI cambia la vita dei super-user? Risultati di uno studio pilota – di <b>Andrea Laudadio</b> , Head of TIM Academy & Development .....	82
Guerra e media: come sinistra e destra raccontano i conflitti in modo diverso – di <b>Edvige Danna</b> , Docente a contratto Unicusano.....	98
Marketing, Intelligenza Artificiale e creatività: uno spartiacque tecnologico – di <b>Massimiliano Moschin</b> , Dottore di ricerca in sociologia e internet studies, IUSVE .....	104
Serious game e AI: il futuro dell'apprendimento è immersivo e personalizzato - di <b>Luca Cavone</b> , Università di Pavia .....	110
Cyberstalking, perché si odia online: dinamiche e ruolo delle piattaforme – di <b>Iva Marino</b> , PhD in Health Promotion and Cognitive Sciences e <b>Francesca Rizzuto</b> , professoressa associata nel Dipartimento Culture e Società dell’Università di Palermo .....	116
Competenze centauriche: L’evoluzione delle competenze e abilità umane nell’era dell’Intelligenza Artificiale Generativa di <b>Andrea Laudadio</b> , Head of TIM Academy & Development .....	126
<b>Errore. Il segnalibro non è definito.</b>	
<b>Sezione DIRITTO DEL DIGITALE .....</b>	<b>145</b>
Cybersicurezza e competenze digitali, pilastri dei diritti – di <b>Carlo Colapietro</b> , Professore ordinario di diritto costituzionale e pubblico presso il dipartimento di giurisprudenza dell’università degli studi Roma Tre .....	147

<i>Allucinazioni dell'intelligenza artificiale: sintomi e cura</i> – di <b>Giovanna Pistorio</b> , Professoressa associata in Diritto costituzionale e pubblico, presso il Dipartimento di Giurisprudenza dell'Università degli Studi Roma Tre.....	162
Il GDPR nell'era di <i>Big Data</i> e IA. Riflessioni su alcune oscillazioni tra obsolescenza e lungimiranza - di <b>Teresa Baldazzi</b> Dottoranda in Law&Social Change presso il Dipartimento di Giurisprudenza dell'Università degli Studi Roma Tre. Esperta con orientamento in Diritto delle nuove tecnologie presso l'Agenzia per la cybersicurezza nazionale.....	173
Le sfide regolatorie della videosorveglianza diffusa tra innovazione tecnologica e tutela dei diritti fondamentali - di <b>Marco Ladu</b> , Professore associato di diritto costituzionale e pubblico presso l'Università degli Studi e-Campus .....	190
Diritto di accesso ad Internet e prospettive d'uso del Metaverso in Sanità – di <b>Lorenzo Maria Lucarelli Tonini</b> , Assegnista di Ricerca Università degli Studi Roma Tre, Dipartimento di Giurisprudenza .....	200
La rete europea dei Garanti per la protezione dei dati: struttura, funzioni e criticità del sistema di governance del GDPR – di <b>Veronica Montozzi</b> , dottoranda di Ricerca in Discipline Giuridiche Pubblicistiche, Internazionalistiche ed Europee presso il Dipartimento di Giurisprudenza dell'Università degli Studi di Roma Tre .....	209
La sanità digitale e le sfide della coesione territoriale e dell'interoperabilità dei sistemi digitalizzati – di <b>Beatrice Perego</b> Dottoressa di ricerca presso l'Università degli Studi Roma Tre, dipartimento di Scienze Politiche .....	225

## Gli LLM pensano? Come funziona davvero il “cervello” digitale dell’IA - di Riccardo Manzotti, Ordinario di Filosofia Teoretica, IULM, Milano

*Il dibattito sull’intelligenza artificiale divide esperti e filosofi. Mentre alcuni negano capacità cognitive alle macchine, altri sostengono che i meccanismi degli LLM riproducano forme genuine di elaborazione intelligente del linguaggio e della conoscenza*

Da quando, nel **2022**, gli LLM (modelli di linguaggio di grandi dimensioni) hanno conquistato il dominio linguistico pubblico, un dibattito—spesso molto acceso—non ha cessato di infuriare: gli LLM sono una forma di intelligenza? **Gli LLM pensano?**

A un estremo vi sono coloro che sostengono che le macchine, pur facendo tantissime cose che se fossero fatte da un essere umano sarebbero un chiaro segno di intelligenza, non sono altro che un’imitazione realizzata attraverso processi matematici. In questo campo, curiosamente, si incontra un’improbabile alleanza. Filosofi, umanisti e studiosi della mente rivendicano uno speciale status ontologico dell’io umano. Allo stesso tempo, vi è un nutrito gruppo di informatici ed esperti di IA—pittorescamente definiti next-tokenisti o parrottisti dal celebre articolo che ha introdotto il termine pappagallo stocastico o *parrot* (Bender et al. 2021)—che, forti di una conoscenza di prima mano dei meccanismi al cuore degli LLM, sostengono che questi modelli si limitano a indovinare la parola successiva senza alcuna reale intelligenza. Insieme, questi autori, sostengono che gli LLM non sono veramente intelligenti, a prescindere da quello che fanno, perché la vera intelligenza, il vero pensiero, è un’altra cosa che solo gli esseri umani sanno fare.

Per certi versi, la loro è una forma di benaltrismo. L’IA vince a scacchi? Il pensiero è ben altro! L’IA traduce come un essere umano? Il pensiero è ben altro! L’IA trova i collegamenti tra l’imperativo categorico di Kant e il principio antropico di John Barrow? Ben altro! Ben altro!

Nel caso dell’IA abbiamo assistito a un classico caso di bersaglio mobile, come hanno recentemente sostenuto **Azeem Azhar e Nathan Warren** sul *Washington Post* (Azhar and Warren 2025). Tutti i criteri che, nell’essere umano e negli animali sono considerati prove di intelligenza, nel momento in cui vengono raggiunti e superati da una macchina, vengono dichiarati insufficienti. Il caso classico è il test di Turing che è stato superato per la prima volta nel 2014 (in era pre-LLM) da un Chatbot che pretendeva di essere un ragazzino ucraino e che non fu riconosciuto come un’IA dal 33% dei giudici umani per oltre 5 minuti di conversazione. Immediatamente, il noto esperto Gary Marcus dichiarò che «non si trattava di vera intelligenza». In seguito, **ChatGPT 4.5 ha passato con margini molto maggiori lo stesso test** risultando indistinguibile da un essere umano. Ogni volta però il test viene dichiarato nullo. Finché era impossibile per una macchina, sembrava una soglia significativa, ma quando viene superato, ecco che non vale più nulla.

In questo articolo vorrei smontare questa posizione, non tanto per dimostrare che gli LLM pensano come gli esseri umani o che il percorso che giunge a una certa proposizione in un LLM o in un umano è uguale. Sarei assolutamente fuori strada. Tuttavia, voglio chiarire che le motivazioni informatico-scientifiche utilizzate dagli scettici sono infondate e che l’intelligenza non è una qualità esclusiva degli esseri umani, ma qualcosa di più generale che si manifesta in situazioni e sistemi non biologici. Non

considererò qui le motivazioni dei filosofi che presuppongono un ruolo metafisicamente privilegiato per gli esseri umani, ma soltanto i ragionamenti di chi—informatici ed esperti di IA—sostengono ci siano meccanismi speciali per «pensare veramente».

Prima di tutto, l'obiezione fondamentale è che non esiste un pensiero interno. Se due sistemi producono lo stesso tipo di risposta, è inutile andare a guardare *dentro* per vedere se uno c'è arrivato pensando e l'altro c'è arrivato facendo calcoli. Quello che conta è l'esito finale. Ovviamente, se l'esito è diverso—mettiamo che la macchina non sappia applicare certi ragionamenti di buon senso—allora ci saranno delle differenze. Ma lo stesso allora si dovrebbe dire di un essere umano che, per vari motivi, non riesca a risolvere particolari problemi. **Il pensiero non è un'essenza invisibile all'interno di un sistema cognitivo, ma è un modo per descrivere la capacità adattiva di certi sistemi** e, nel caso del linguaggio, la capacità di utilizzare in modo generativo la conoscenza. Non ci sono processi che producono il pensiero all'interno di un sistema cognitivo, semplicemente perché il pensiero in questo senso ingenuo non esiste. Tra i neuroni non ci sono pensieri. I neuroni fanno quello che devono fare, organizzando stimoli in entrata e uscita e producendo il giusto comportamento, anche attraverso stati intermedi che mediano tra stimoli lontani e prossimi e pilotano comportamenti complessi. In mezzo però non c'è una cosa aggiuntiva e invisibile che chiamiamo pensiero.

Entrando nel merito dei meccanismi, gli scettici (il mio termine favorito è *next-tokenisti*) fanno un errore concettuale ed empirico comparando il *meccanismo di base* dell'LLM (la probabilità condizionata) con il *comportamento globale* dell'essere umano. **Questo è fuorviante**. Sarebbe altrettanto sbagliato comparare il meccanismo di base del cervello (apprendimento hebbiano e soglia non lineare dei neuroni) con il comportamento globale dell'LLM. Si concluderebbe simmetricamente che noi siamo meno intelligenti di un LLM, perché un neurone non sa esprimersi linguisticamente. Sarebbe una conclusione assurda.

Il ragionamento corretto dovrebbe confrontare i meccanismi di base dell'LLM con quelli dell'umano, o confrontare i comportamenti globali. Il singolo neurone non è affatto più intelligente, intenzionale, cosciente, mentale, volitivo del singolo tensore di una LLM. La comprensione del testo non avviene a livello dei componenti di base della rete neurale, ma è frutto della dinamica globale tra la rete neurale e l'ambiente. Un LLM a *livello di dinamica globale con l'ambiente* (il dataset di addestramento e i prompt con gli esseri umani) esibisce un comportamento linguistico che—al netto di affermazioni metafisiche—permette di confrontare esseri umani e IA generative. Oltre tutto in questa prospettiva neutra, recentemente modellata dal cosiddetto sistema zero (Chiriatti et al. 2024), è interessante considerare una progressiva integrazione tra intelligenza umana e macchinica. È un punto di vista che, in passato, era stato espresso anche dal principio di parità cognitiva secondo cui il modo in cui un processo cognitivo è implementato non è discriminante; tesi che è uno sviluppo del funzionalismo (Clark and Chalmers 1998; Putnam 1960).

Ma non è finita qui. Anche considerando in maggiore dettaglio il funzionamento interno di un LLM non ci sono motivi per negargli cittadinanza cognitiva, ovviamente al netto del fatto che, fino ad oggi, l'IA generativa non ha potuto disporre di un corpo e quindi non è stata incarnata al pari di un agente biologico. Consideriamo il funzionamento di un LLM; al suo cuore c'è **l'algoritmo Transformer** in qualche sua recente incarnazione (Vaswani et al. 2017; Bubeck et al. 2023; OpenAI 2023; Agashe et al. 2024; Jin et al. 2024; Mangalam 2025). Quelli che sostengono che gli LLM non fanno altro che indovinare la parola successiva, semplificano eccessivamente il loro meccanismo interno, limitandosi a fare un corto circuito tra il meccanismo di base (ovvero la statistica bayesiana) e il comportamento complessivo del sistema (ovvero la generazione di un flusso di token a partire da un prompt).

Questo modo di ragionare mostra tutti i suoi limiti se applicato al cervello. In fondo, si potrebbe dire, un neurone non è altro che una sommatoria a soglia non lineare con apprendimento hebbiano. Globalmente, però, il cervello esibisce un comportamento linguistico che qualcuno potrebbe ridurre a mettere una parola dietro l'altra: cioè quello che sto facendo io in questo momento per scrivere questo articolo. Il punto è che, per arrivare a prevedere quella famosa “parola successiva” o *next token*—tra il basso livello (un tensore bayesiano o una sinapsi hebbiana) e l’output finale—deve succedere qualcosa in mezzo ed è qui che l’argomento del pappagallo stocastico perde la sua forza.

Entriamo ancora di più nel dettaglio dell’algoritmo Transformer tenendo conto che, tra l’idea di base e l’implementazione commerciale di un LLM, ci sono numerosi livelli che aggiungono ulteriori gusci di complessità. Partiamo dal cuore del sistema, l’equazione dell’attenzione (Vaswani et al. 2017):

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{D_k}}\right)V$$

In termini qualitativi, lo scopo di questa equazione ( $D_k$  è solo un fattore numerico di normalizzazione), è il calcolo delle matrici dell’attenzione  $Q$  e  $K$  sulla base del dataset  $V$ . Praticamente,  $V$  contiene una porzione di testo su cui fare il training e  $Q$  e  $K$  sono le matrici che devono essere modificate fino a essere in grado di autogenerare  $V$  anche quando  $V$  è parzialmente nascosto. A ogni iterazione, si toglie un pezzetto (un token o più) di  $V$  e si prova a modificare  $Q$  e  $K$  in modo che rigenerino la totalità di  $V$ . In questo modo  $Q$  e  $K$  estraggono la conoscenza (e non soltanto la probabilità condizionale) contenuta nei testi usati per fare il training. In questo modo, quando il sistema riceve un prompt, genera il testo che, sulla base di tutta la conoscenza contenuta nei testi prodotti fino a quel momento, gli autori umani di questi testi, avrebbe probabilmente prodotto.  $Q$  e  $K$  costruiscono un modello statistico che incarna la conoscenza che nei testi si è tradotta in legami probabilistici tra le parole. In questo modo  $Q$  e  $K$  diventano in grado di generare nuovi testi che riflettono la struttura sintattico-semantica e, perché no, epistemica del testo di training.

Ma non finisce qui. **Un LLM non si limita a cercare legami tra parole, ma anche tra strutture più astratte e più globali.** I vari ChatGPT, Grok, Claude, DeepSeek o Gemini applicano l’equazione dell’attenzione sfruttando una struttura gerarchica iterativa in grado di creare collegamenti non solo tra parole (il famoso *next token*), ma anche tra strutture bayesiane a molteplici livelli di astrazione. Ovviamente, a livello astratto, causalità e probabilità condizionata sono cose diverse, ma è un dato di fatto che la statistica bayesiana è una delle migliori approssimazioni matematiche che si possono dare della causalità tra fenomeni. Fatte le debite ipotesi, un valore di  $P(A|B)$  pari a 1 è un ottimo indicatore del fatto che  $B$  causi  $A$ .

L’amplificazione dell’attenzione a livelli gerarchici multipli è implementato dal cosiddetto *multi-head* già introdotto da Vaswani (Vaswani et al. 2017), ma poi successivamente migliorato (Kwiatkowski et al. 2019; Granier and Senn 2025). Il *multi-head* consiste nel procedere per vie multiple (in senso massivo) all’interno del dataset. Ogni *head* (“testa”) è un meccanismo di attenzione indipendente, cioè un piccolo blocco che prende l’input, lo proietta in uno spazio ridotto, calcola le matrici  $Q$  e  $K$  in quel contesto e produce un output. Il dataset viene così percorso a livelli diversi di astrazione e di complementarietà. In questo modo si estraggono relazioni sintattiche e semantiche a livelli molteplici di astrazione. Non è conoscenza linguistica questa? Non è, a tutti gli effetti, **l’implementazione del livello epistemico e semiotico?** A parità di altre condizioni, non è, in fondo, quello che facciamo noi esseri umani quando leggiamo e costruiamo una rete di relazioni tra “concetti”? Che cosa sono i

concetti se non strutture condizionali nell'immensa rete di input linguistici che riceviamo fin da piccoli leggendo e ascoltando?

Gli scettici, a questo punto, lamentano che non ci sono strutture che corrispondano a meccanismi deduttivi, pensiero o altro. Per esempio, Luciano Floridi ha ripetutamente dichiarato che l'IA avrebbe «intelligenza zero» e secondo lui è una sciocchezza che credere che l'IA «possa diventare intelligente anche solo come un gatto, che capisca o possa un giorno capire».

In altre circostanze, **Walter Quattrociocchi ha sostenuto che gli LLM non sono intelligenti, perché non capiscono, ma fanno statistica**. Per lui, gli «LLM non capiscono il mondo: predicono la parola successiva in una sequenza, sulla base di correlazioni osservate in enormi corpus testuali. È un'operazione statistica, non epistemica. L'apparenza di intelligenza nasce dalla nostra propensione a riconoscere significato e intenzionalità in ciò che ha forma linguistica.» (Quattrociocchi 2025). È la posizione perfetta dei next-tokenisti! Non trovo niente di sbagliato nella sua analisi degli LLM.

Quattrociocchi non ha torto per quanto riguarda il meccanismo bayesiano, ma è sicuro che il cervello faccia qualcosa di diverso? È sicuro che un sistema di milioni di parametri che catturano le relazioni epistemiche multiple come l'attenzione e il multi-head non catturi la conoscenza dal linguaggio?

Il pensiero e il capire, per quanto è dato da vedere, non esistono neppure a livello neurale se è per questo (Manzotti and Rossi 2023). Ovviamente, un LLM manca di alcuni tasselli fondamentali per riprodurre una mente, tra i quali direi il corpo, la coscienza e le intenzioni. Ma l'intelligenza come *leggere tra*, come meccanismo di estrazione della conoscenza epistemica da un corpus linguistico? Non mi sembra manchi nulla. Il meccanismo multi-head applicato all'attenzione su scale multiple dell'algoritmo Transformer fa proprio questo: estrae dal linguaggio tutta la struttura causale-condizionale. Questa struttura condizionale è molto vicina a quello che potremmo definire il livello semiotico dell'intelligenza. Non è un caso che l'etimologia del termine «intelligenza» derivi da «leggere tra» ovvero cogliere le relazioni tra i termini del linguaggio o di un problema piuttosto che arrivate mentalmente al significato del mondo attraverso una fantomatica capacità, come la semantica o l'intenzionalità.

Quando gli esperti negano che oggi gli LLM pensino veramente, una domanda che sorge spontanea è «che cosa intendono per pensare veramente?». Che cosa sarebbe la vera intelligenza? Molti di loro hanno ereditato dalle scienze cognitive delle rappresentazioni del nostro pensiero (per esempio in termini di concetti e deduzioni) e sono alla ricerca di queste rappresentazioni all'interno dei meccanismi. Non si rendono conto di confondere il modello dell'intelligenza con l'intelligenza stessa. C'è un'autorevole tradizione che trovava nella competenza linguistica la base dell'intelligenza e cioè del *leggere tra* (Benveniste 1971). Ovviamente, un'animale o un essere umano hanno molto di più. Come si era detto sopra hanno un corpo, sono coscienti e hanno obiettivi. Ma l'intelligenza? Che cosa sarebbe di più? I next tokenisti conoscono bene i meccanismi alla base degli LLM, ma cosa sanno del pensiero e del capire? Forse, al netto del contorno ambientale e corporeo, non c'è veramente niente altro. Non esiste nessuna definizione fisica del pensiero in quanto presenza invisibile tanto nei processi cognitivi quanto nell'attività neurale.

In fondo, già un secolo fa il grande **Ludwig Wittgenstein** ammoniva che «una delle idee logicamente più pericolose è l'idea che pensiamo con la testa, o nella testa. [...] L'idea del pensare come di un processo che ha luogo nella testa; in uno spazio perfettamente chiuso, conferisce al pensare un che di occulto.» (Wittgenstein 2007). È strano che oggi molti esperti di IA difendano una posizione che ha qualcosa di magico.

## References

- Agashe, Saaket, Jiuzhou Han, Shuyu Gan, Jiachen Yang, Ang Li, and Xin Eric Wang. 2024. 'Agent S: An Open Agentic Framework that Uses Computers like a Human', *arXiv*, 2410.08164v1.
- Azhar, Azeem, and Nathan Warren. 2025. 'The paradox of GPT-5', Accessed 14/8/2025.
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. "On the Dangers of Stochastic Parrots." In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610-23.
- Benveniste, E. 1971. *Problemi di linguistica generale* (Il saggiaore: Milano).
- Bubeck, Sebastian, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Hortvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yanzhi Li, Li Scott, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio, and Yi Zhang. 2023. 'Sparks of Artificial General Intelligence: Early experiments with GPT-4', *arXiv*, March 22: 1-154.
- Chiriatti, Massimo, M. Ganapini, E. Panai, M. Ubiali, and Giuseppe Riva. 2024. 'The case for human-AI interaction as system 0 thinking', *Nature Human Behavior*, 8: 1829-30.
- Clark, Andy, and David John Chalmers. 1998. 'The Extended Mind', *Analysis*, 58: 10-23.
- Granier, Arno, and Walter Senn. 2025. 'Multihead self-attention in cortico-thalamic circuits', *arxiv/Physics*.
- Jin, Peng, Bo Zhu, Li Yuan, and Shuicheng Yan. 2024. 'MOH: Multi-head Attention as Mixture-of-Head Attention', *arxiv/Physics*, 2410.11842v1
- Kwiatkowski, Tom, Jennimaria Palomaki, Olivia Redfield, Michael Collins, Ankur Parikh, Chris Alberti, Danielle Epstein, Illia Polosukhin, Jacob Devlin, and Kenton Lee. 2019. 'Natural questions: a benchmark for question answering research', *Transactions of the Association for Computational Linguistics*, 7: 453-66.
- Mangalam, Madhur. 2025. 'The Variational Necromancy: A Forensic Autopsy of the Free Energy Principleis Academic Corpse', *arXiv*, May 28.
- Manzotti, Riccardo, and Simone Rossi. 2023. *IO & IA. Mente, cervello & GPT* (Rubbettino: Soveria Mannelli).
- OpenAI. 2023. 'GPT-4 Technical Report', *arXiv*, March 2023.
- Putnam, Hilary. 1960. 'Minds and Machines.' in Sidney Hook (ed.), *Dimensions of Mind* (New York University Press: New York).
- Quattrociocchi, Walter. 2025. 'L'AI non è davvero intelligente, è una questione di statistica», *Corriere della sera*.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lucasz Kaiser, and Lilia Polosukhin. 2017. 'Attention Is All You Need', *arXiv*: 1706.03762v5.
- Wittgenstein, Ludwig. 2007. *Zettel* (Einaudi: Torino).

## **Manipolazione e inganno nei robot sociali: i paletti dell’AI Act – di Andrea Bertolini, Scuola Superiore Sant’Anna di Pisa, Rocco Limongelli, Scuola Superiore Sant’Anna di Pisa, Alessandra Rossi, Dipartimento di ingegneria elettrica e tecnologie dell’informazione, Università di Napoli, Federico II e Silvia Rossi, Università di Napoli Federico II**

*I robot sociali trovano spazio in settori diversi, ma l’AI Act stabilisce regole stringenti sulle pratiche ingannevoli e manipolative, introducendo incertezze interpretative sul concetto di danno significativo, con effetti diretti su ricerca scientifica, diffusione tecnologica e industria della robotica*

I progressi nei campi dell’intelligenza artificiale (IA), dell’apprendimento automatico, della robotica e dell’interazione uomo-robot (HRI) stanno favorendo lo sviluppo e la diffusione dei robot sociali per un’interazione con gli esseri umani in modo socialmente accettabile e più naturale. I robot sociali (SR) sono generalmente definiti come robot interattivi che «esibiscono le seguenti caratteristiche: esprimono e/o percepiscono emozioni; comunicano tramite dialoghi di alto livello; apprendono modelli o riconoscono altri agenti; instaurano e/o mantengono relazioni sociali; utilizzano segnali naturali (sguardo, gesti, ecc.); mostrano una personalità e un carattere distintivi; e possono apprendere e/o sviluppare competenze sociali» [1].

Il loro scopo principale è coinvolgere gli utenti umani e stabilire connessioni con loro per raggiungere un ampio spettro di obiettivi, che varia da l’intrattenimento [2], alla compagnia [3], alla fornitura di cure [4], e a un certo livello di assistenza [5]. Ad esempio, nel settore educativo, robot come quelli sviluppati nell’ambito del progetto L2TOR hanno dimostrato efficacia nel migliorare sia gli esiti cognitivi sia quelli affettivi, con risultati paragonabili a quelli ottenuti da tutor umani [6]. I robot sociali trovano inoltre ampia applicazione nei settori del commercio al dettaglio e dell’ospitalità, offrendo assistenza ai clienti 24 ore su 24; sono impiegati per rispondere alle domande, guidare i clienti, fornire raccomandazioni personalizzate [7]. Nell’ambito dell’assistenza, la loro applicazione spazia dalla cura degli anziani [8], un’esigenza urgente vista la crescente popolazione anziana a livello globale [9], fino alla terapia cognitivo-comportamentale per bambini con disturbo dello spettro autistico. In scenari assistivi, i robot sociali sono usati anche come supporto per il cambiamento comportamentale delle persone (*behavioural change*) contribuendo a migliorare significativamente la qualità della vita di anziani e pazienti [10].

Un **robot sociale che mira a influenzare i comportamenti deve interagire in modo strategico, tenendo conto del contesto e dello stato psicologico e benessere delle persone**, anche ricorrendo a forme di inganno, come il fuorviare sugli stati contestuali o le intenzioni altrui, per ottenere effetti prosociali positivi [11]. L’uso dell’inganno nell’interazione uomo-robot può prevenire conflitti, ridurre lo stress emotivo e migliorare le relazioni, attraverso diverse tipologie di interazione (ad esempio, piccole bugie o errori volontari) che si sono dimostrate utili in ambito educativo, sanitario e assistenziale [12]. C’è da notare, però, che tali strategie di interazione possono avere effetti sia positivi che negativi sugli utenti [13]. Pur migliorando i servizi e le applicazioni assistive grazie a esperienze personalizzate e a una

maggiori efficienze operativa, i robot sociali devono raccogliere e processare grandi quantità di dati per comprendere dinamiche umane complesse e generare risposte adeguate. Questa raccolta dati può così comportare conseguenze negative sia nelle dinamiche uomo-robot sia a livello individuale, poiché una violazione della privacy può ridurre l'accettazione, la fiducia e la disponibilità all'uso del robot, oltre a sollevare questioni etiche e legali.

Queste applicazioni dei robot sociali hanno fatto emergere la necessità di bilanciare i loro benefici con le potenziali implicazioni etiche, di privacy e sicurezza. A tal fine, la Commissione Europea (CE) ha istituito un quadro normativo comune (il cosiddetto AI Act) per proteggere gli individui dagli effetti negativi derivanti dall'applicazione dei sistemi di intelligenza artificiale. Sebbene l'AI Act non impatti direttamente la ricerca nei campi dell'IA e dell'interazione uomo-robot, esso ha introdotto limiti poco chiari sull'uso di tecniche manipolative, persuasive e ingannevoli nei sistemi dotati di intelligenza artificiale. La parte principale dell'AI Act (AIA) riguarda le disposizioni applicabili solo ai cosiddetti sistemi di intelligenza artificiale ad alto rischio (H-AIS), definiti all'Articolo 6. I sistemi non H-AIS sono invece soggetti a requisiti meno rigorosi, principalmente legati alla trasparenza. Tuttavia, alcune applicazioni dell'IA sono considerate illecite e vietate esplicitamente dall'Articolo 5 dell'AIA. Alcune di queste pratiche di IA sono descritte in modo così ampio che le applicazioni dei robot sociali potrebbero ricadere tra queste, rischiando quindi di essere proibite. Sebbene la ricerca sia esentata, la distribuzione sul mercato con tali pratiche vietate diventa illecita. Se i robot sociali venissero classificati come illeciti, ciò potrebbe compromettere radicalmente il loro sviluppo e la loro diffusione sul mercato, mettendo a rischio la crescita di questo specifico settore della robotica.

In un mondo in cui l'intelligenza artificiale è ormai parte integrante della vita quotidiana, influenzando le scelte e i comportamenti delle persone, dai sistemi di raccomandazione di contenuti streaming fino agli esiti elettorali (come nel caso della controversia Cambridge Analytica), **l'Artificial Intelligence Act (AI Act)** del 21 aprile 2021 rappresenta un tentativo di creare un quadro regolatorio che sia al contempo stimolante e restrittivo. La Commissione Europea sottolinea, nella memoria esplicativa che accompagna l'AI Act, il duplice obiettivo del regolamento: da un lato promuovere un uso responsabile e migliorato dell'IA; dall'altro identificare e mitigare i rischi potenziali associati a questa tecnologia, adottando un approccio progettuale basato sulla sicurezza e sull'attenzione all'essere umano, con l'intento di tutelare i diritti fondamentali. L'AI Act rappresenta così l'equilibrio dell'Unione Europea tra la protezione dei diritti individuali e la promozione dell'innovazione tecnologica, ponendo al centro la necessità di governare la crescita digitale in un mondo sempre più guidato dagli algoritmi, verso un futuro più sicuro e trasparente abilitato dall'intelligenza artificiale.

L'AI Act rappresenterà una pietra miliare, stabilendo la prima definizione giuridicamente vincolante di intelligenza artificiale nell'Unione Europea. La definizione di IA delineata dall'AI Act all'articolo 3, paragrafo 1, recita: «*Un sistema di intelligenza artificiale è un sistema basato su macchine progettato per operare con diversi livelli di autonomia e che può mostrare adattabilità dopo la messa in servizio e che, per obiettivi esplicativi o impliciti, deduce, dall'input che riceve, come generare output quali previsioni, contenuti, raccomandazioni o decisioni in grado di influenzare ambienti fisici o virtuali.*».

Sebbene l'AI Act fornisca una definizione completa di intelligenza artificiale e preveda un ambito regolatorio ampio, riconosce anche la necessità di specifiche eccezioni. Questo consente al regolamento di concentrarsi sulle aree in cui regole armonizzate possono avere il massimo impatto, evitando di sovrapporsi a normative esistenti che già affrontano questioni legate all'IA.

**L'AI Act dell'UE si applicherà anche ai robot**, intesi come sistemi di intelligenza artificiale integrati in dispositivi hardware. Nel Considerando 28 dell'AI Act si afferma che: «*I sistemi di IA potrebbero causare effetti negativi sulla salute e sicurezza delle persone, soprattutto quando operano come componenti di prodotti. In linea con gli obiettivi della normativa dell'Unione per armonizzare il mercato interno e garantire che solo prodotti sicuri e conformi vengano immessi sul mercato, è importante prevenire e mitigare i rischi di sicurezza generati dal prodotto nel suo complesso, inclusi quelli derivanti dai componenti digitali come i sistemi di IA. Ad esempio, robot sempre più autonomi, sia nel contesto della produzione che nell'assistenza personale e sanitaria, devono poter operare in sicurezza e svolgere le loro funzioni in ambienti complessi.*».

Un altro riferimento ai robot si trova nella Comunicazione della Commissione Europea che definisce l'IA come: «*Sistemi che mostrano comportamenti intelligenti analizzando l'ambiente e prendendo decisioni con un certo grado di autonomia per raggiungere obiettivi specifici. I sistemi basati su IA possono essere puramente software, operando nel mondo virtuale (es. assistenti vocali, software di analisi immagini, motori di ricerca, sistemi di riconoscimento vocale e facciale), oppure l'IA può essere integrata in dispositivi hardware (es. robot avanzati, auto autonome, droni o applicazioni Internet delle cose)*».

Tuttavia, la Commissione Europea non fornisce una definizione esaustiva dei robot all'interno dell'AI Act. Questo approccio si rivela problematico e ci porta inevitabilmente alla domanda centrale di questo lavoro: «*dove si collocano i robot sociali all'interno dell'AI Act?*»

**L'art. 5, §1 dell'AIA** stabilisce che sono vietate le seguenti pratiche di intelligenza artificiale:

- la messa in commercio, l'immissione in servizio o l'uso di un sistema di IA che **impieghi tecniche subliminali al di fuori della consapevolezza della persona o tecniche manipolative o ingannevoli intenzionali**, con l'obiettivo o l'effetto di alterare in modo significativo il comportamento di una persona o di un gruppo di persone, **compromettendo apprezzabilmente la loro capacità di prendere decisioni informate**, causando così che esse prendano una decisione che altrimenti non avrebbero preso, in modo tale da causare o essere ragionevolmente probabile che causi **danni significativi** a quella persona, a un'altra persona o a un gruppo di persone;
- la messa in commercio, l'immissione in servizio o l'uso di un sistema di IA che **sfrutti qualsiasi vulnerabilità di una persona fisica o di un gruppo specifico di persone** dovuta alla loro età, disabilità o a una particolare situazione sociale o economica, con l'**obiettivo o l'effetto di alterare materialmente il comportamento** di quella persona o di una persona appartenente a quel gruppo in modo tale da **causare o essere ragionevolmente probabile che causi danni significativi** a quella persona o ad un'altra persona.

Questa citazione evidenzia il tipo di applicazioni che rientrano nel divieto, rappresentando gli aspetti principali che i ricercatori nel campo della robotica e dell'interazione uomo-robot devono definire per poter implementare i robot sociali assistivi in applicazioni reali. Riteniamo

che particolare attenzione debba essere posta ai seguenti aspetti:

- **Tecniche subliminali al di fuori della consapevolezza della persona:** Le interazioni dei robot sociali mirano a essere percepite come naturali e confortevoli, e poiché piccoli cambiamenti spesso sfuggono alla consapevolezza degli utenti [14], solo modifiche significative nel comportamento — fisico, cognitivo o sociale — risultano efficaci per favorire adattamento, personalizzazione e accettazione tecnologica [15].
- **Tecniche manipolative intenzionali:** Le applicazioni di robot socialmente intelligenti, pensate soprattutto per persone fragili, mirano a stimolare il coinvolgimento e creare legami personali attraverso interazioni socio-emotive complesse, design mirato e adattamento nel tempo, risultando particolarmente efficaci per utenti con difficoltà di socializzazione [16, 17, 18].
- **Tecniche ingannevoli:** Nella robotica sociale l'inganno, pur sollevando questioni etiche e giuridiche [19], è spesso incorporato nel design come strumento per favorire accettazione e coinvolgimento degli utenti, rendendo però difficile distinguere tra inganno vero e proprio e soluzioni funzionali all'interazione.
- **Alterazione materiale del comportamento:** L'impatto dei robot sociali sul comportamento umano può essere positivo o negativo a seconda degli esiti prodotti, ma l'assenza di criteri chiari per valutarne la desiderabilità solleva questioni etiche e giuridiche legate alla definizione di "danno significativo" e alla tutela degli utenti [20].
- **Sfruttamento delle vulnerabilità:** I robot sociali sono progettati per gruppi vulnerabili come anziani, bambini e persone con disabilità, ma la desiderabilità del loro impiego dipende dagli esiti, che variano tra individui e categorie, rendendo complessa la valutazione e problematica l'idea che solo alcuni siano vulnerabili.

Tutti gli aspetti sopra descritti sono particolarmente rilevanti per la robotica sociale perché hanno in comune il concetto di "**Significant Harm**", danno significativo. L'art. 5 consente applicazioni manipolative o ingannevoli rivolte a vulnerabilità degli utenti, purché non provochino un "danno significativo", concetto giuridico centrale ma affidato alla discrezionalità dei giudici e al bilanciamento tra interessi contrapposti. Nel diritto civile, il danno comprende sia perdite patrimoniali che non patrimoniali, ma nel contesto della robotica sociale i pregiudizi psicologici o relazionali sono difficili da dimostrare per l'incertezza del nesso causale. La valutazione del "danno significativo" deve quindi essere contestuale e sistematica, distinguendo tra danni individuali risarcibili e impatti collettivi che possono giustificare divieti.

La nozione di "danno significativo" sembra lasciare un eccessivo margine di incertezza e di valutazione discrezionale *ex post* da parte di singoli interpreti.

Sebbene sia certo che i decisori politici europei non intendessero vietare radicalmente tutte le applicazioni della robotica sociale, ostacolando così lo sviluppo di uno dei settori più promettenti della robotica europea, è altrettanto certo che le disposizioni contenute nell'Art. 5, §1, lettere (a) e (b) dell'AI Act possano colpire alcune applicazioni della robotica sociale. Un approccio centrato sull'essere umano allo sviluppo tecnologico — come quello promosso dalla Commissione Europea, a partire dalla sua comunicazione su un approccio europeo all'IA — richiede che le tecnologie si conformino ai principi e ai valori essenziali riflessi nella regolamentazione europea. Le disposizioni dell'Art. 5 dell'AIA mirano a garantire il massimo

livello di tutela e rispetto dell'individuo, e sotto questo profilo vanno certamente apprezzate.

L'analisi del AI Act e delle sue possibili ricadute sulla ricerca in robotica sociale richiede di affrontare i seguenti aspetti, che meritano una discussione approfondita:

- **I criteri usati dalle norme** (Art. 5, §1, lettere (a) e (b)) per identificare le pratiche di IA proibite non sono sufficientemente definiti, né di natura puramente giuridica. Essi richiedono piuttosto un approccio multidisciplinare che coinvolga psicologia, ingegneria e filosofia, e che può portare a risultati differenti a seconda delle specifiche definizioni adottate per ciascuno di questi concetti. I criteri individuati possono essere interpretati in molteplici modi e basarsi sulla valutazione di diversi elementi per essere considerati soddisfatti. Particolarmente rilevante è la definizione di manipolazione e inganno, e la modalità con cui si determina il targeting di categorie potenzialmente fragili di individui;
- **Sussiste un'incertezza significativa riguardo a cosa la ricerca e lo sviluppo di robot sociali possano considerare pienamente leciti e ammissibili.** Molte applicazioni utilizzano intenzionalmente qualche forma di manipolazione e inganno degli utenti per facilitare l'interazione, che non sempre viene percepita negativamente. In particolare, in tali applicazioni i robot sociali tengono sistematicamente conto delle caratteristiche degli utenti, incluse le loro vulnerabilità, per assicurare un'adeguata personalizzazione. Pertanto, in assenza di una definizione chiara e univoca di tutti i requisiti previsti dalle disposizioni, il criterio probabilmente più rilevante per definire cosa debba essere considerato lecito o meno per queste applicazioni di robot sociali è la nozione di danno significativo;
- **Sebbene il concetto di “danno significativo” sia giuridico, esso è comunque poco definito,** dato che le conseguenze positive e negative derivanti dall'uso e dall'interazione con i robot sociali sono molteplici, impattando principalmente — sebbene non esclusivamente — la dimensione psicologica e relazionale della vita degli utenti. Di conseguenza, è difficile stabilire con precisione quali aspetti debbano essere considerati per valutare se l'impatto di un robot sociale sull'utente costituisca “danno significativo”. Occorre considerare la condizione fisica? L'impatto sulla capacità di stabilire relazioni interpersonali? L'effetto su terze parti (ad esempio i familiari) del trattamento basato sul robot? Dovrebbe essere considerato il danno subito da questi ultimi (ad esempio la riduzione delle interazioni tra utenti e familiari, qualora i primi preferiscano il robot sociale)? Queste sono solo alcune delle problematiche e incertezze che emergono nel bilanciare interessi e fattori opposti che una valutazione di “danno significativo” richiederebbe. Gran parte di questa incertezza potrà essere colmata solo con un approccio multidisciplinare, che includa anche l'etica, lasciando però un ampio margine a decisioni arbitrarie e a incertezza ex-ante;
- Considerando che **la decisione di proibire una certa pratica di IA** riguarda un'intera classe di utenti, oltre ai ricercatori e sviluppatori che hanno investito nel sistema IA, come dovrebbero influire le esperienze di singoli individui e/o di sottogruppi e sottocategorie nel determinare la liceità o meno di un dato robot sociale e del suo utilizzo?
- Considerando che il “danno significativo” potrebbe non essere ancora stato verificato al momento della decisione di proibire una pratica, ma solo anticipato, come dovrebbe essere trattata questa incertezza e come dovrebbe influire sulla decisione degli sviluppatori e ricercatori di investire o meno in una determinata tecnologia?
- Il fatto che diversi interpreti, inclusi giudici nei vari Stati membri europei, possano dover

applicare tali disposizioni in assenza di criteri chiari, univoci e di rilevanza oggettiva, porterà presumibilmente a decisioni divergenti tra le giurisdizioni, con conseguenze che potrebbero colpire profondamente l'industria europea, generando incertezza su quali pratiche di IA siano lecite e quali proibite. Questo porterà a una frammentazione del mercato e a incentivi negativi per investire nello sviluppo di tali tecnologie a livello europeo. In particolare, ciò penalizzerà maggiormente l'industria europea rispetto a quella non europea, considerando la dimensione più contenuta rispetto ai grandi attori globali che potrebbero accettare maggiori livelli di incertezza legale e comunque immettere le tecnologie sul mercato, correndo il rischio di cause legali ex post.

Pur ritenendo che l'AIA debba essere rivista tenendo conto dei suddetti aspetti, consigliamo anche ai ricercatori in robotica sociale di integrare ulteriormente comportamenti trasparenti nei robot [19]. Crediamo che se le funzionalità del sistema sono percepite come chiare e accettabili dall'utente, anche alcune forme di manipolazione della percezione umana potrebbero non essere considerate al di fuori delle funzionalità stesse del sistema. I comportamenti trasparenti dei robot potrebbero mitigare alcuni rischi, permettendo alle persone di identificare anche sottili "stati" o "intenzioni" robotiche [21]. Inoltre, sosteniamo un design centrato sull'utente per queste tecnologie, sviluppandole a partire dai bisogni reali delle popolazioni fragili e vulnerabili. Nel caso dei robot sociali, coinvolgere gli utenti finali e le parti interessate fin dalle prime fasi di progettazione e sviluppo è un modo valido per evitare che tecniche di apprendimento, assistenza e adattamento, sebbene utilizzate per soddisfare bisogni specifici, vadano anche a sfruttare le vulnerabilità delle persone.

In questo articolo abbiamo analizzato e discusso gli effetti dell'AI Act, con particolare attenzione all'art. 5, sullo sviluppo di robot sociali destinati a essere immessi sul mercato e impiegati nel mondo reale. Abbiamo fornito definizioni ed esempi in cui i robot sociali potrebbero essere considerati illeciti e bloccati, anche se la ricerca *allo stato dell'arte* ha dimostrato i benefici della robotica sociale sul benessere della popolazione.

L'analisi presentata sottolinea la fondamentale importanza che ricercatori in robotica, intelligenza artificiale, psicologia, sociologia, filosofia e decisori politici forniscano indicazioni su come i diversi criteri elencati all'art. 5, §1, lett. (a) e (b) debbano essere definiti e interpretati. Poiché linee guida come quelle previste dall'art. 96, §1, lett. (b) dell'AIA non avranno valore giuridico vincolante, i giudici non saranno impediti dal decidere e applicare i criteri esistenti nell'AIA in una molteplicità di modi, il che influenzerà profondamente il mercato e la possibilità per l'industria europea di prosperare. Inoltre, riteniamo che i ricercatori debbano fornire definizioni sufficientemente dettagliate, poiché solo un approccio analitico potrà compensare la mancanza di una stringente efficacia giuridica.

## **Ringraziamenti**

Questo materiale si basa su un lavoro sostenuto dal programma di ricerca e innovazione Horizon 2020 dell'Unione Europea nell'ambito della borsa Marie Skłodowska-Curie, grant agreement n. 955778 (R. Limongelli), (A. Bertolini), e dal Ministero Italiano della Ricerca nell'ambito delle azioni complementari del NRRP Fit4MedRob - Fit for Medical Robotics Grant

No. PNC0000007 (A. Rossi e S. Rossi).

## Bibliografia

1. Dautenhahn Kerstin (2007) Socially intelligent robots: dimensions of human–robot interactionPhil. Trans. R. Soc. B362679–704 <http://doi.org/10.1098/rstb.2006.2004>
2. Nitha Elizabeth John, Alessandra Rossi, and Silvia Rossi. 2022. Personalized Human-Robot Interaction with a Robot Bartender. In Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '22 Adjunct). Association for Computing Machinery, New York, NY, USA, 155–159. <https://doi.org/10.1145/3511047.3537686>
3. Di Napoli, C., Ercolano, G. & Rossi, S. Personalized home-care support for the elderly: a field experience with a social robot at home. *User Model User-Adap Inter* 33, 405–440 (2023). <https://doi.org/10.1007/s11257-022-09333-y>
4. G. Maggi, L. Raggioli, A. Rossi and S. Rossi, "Assessment of Distraction and the Impact on Technology Acceptance of Robot Monitoring Behaviour in Older Adults Care," in IEEE Transactions on Affective Computing, doi: 10.1109/TAFFC.2025.3539015.
5. A. Rossi, G. Falcone and S. Rossi, "It Is the Way You Lie: Effects of Social Robot Deceptions on Trust in an Assistive Robot," 2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN), Pasadena, CA, USA, 2024, pp. 1106-1111, doi: 10.1109/RO-MAN60168.2024.10731172
6. Tony Belpaeme et al., Social robots for education: A review. *Sci. Robot.* 3, eaat5954 (2018). DOI:10.1126/scirobotics.aat5954
7. Rossi, A., Rossi, S., Maro, M.D. et al. BRILLO: Personalised HRI with a Bartender Robot. *Int J of Soc Robotics* (2025). <https://doi.org/10.1007/s12369-025-01228-3>
8. Broadbent, E., Stafford, R. & MacDonald, B. Acceptance of Healthcare Robots for the Older Population: Review and Future Directions. *Int J of Soc Robotics* 1, 319–330 (2009). <https://doi.org/10.1007/s12369-009-0030-6>
9. Zlotnik, Hania, Population projections and world population trends, Demography-VOLUME II, 314, 2010, EOLSS Publications
10. Dhana K, Franco OH, Ritz EM, Ford CN, Desai P, Krueger KR, Holland TM, Dhana A, Liu X, Aggarwal NT, Evans DA, Rajan KB. Healthy lifestyle and life expectancy with and without Alzheimer's dementia: population based cohort study. *BMJ*. 2022 Apr 13;377:e068390. doi: 10.1136/bmj-2021-068390
11. Sætra HS. The foundations of a policy for the use of social robots in care. *Technol Soc.* 2020 Nov;63:101383. doi: 10.1016/j.techsoc.2020.101383
12. R. C. Arkin, P. Ulam and A. R. Wagner, "Moral Decision Making in Autonomous Systems: Enforcement, Moral Emotions, Dignity, Trust, and Deception," in Proceedings of the IEEE, vol. 100, no. 3, pp. 571-589, March 2012, doi: 10.1109/JPROC.2011.2173265.
13. Robot Sociali per l'Assistenza: Inganni Pro-sociali ed Etica / Rossi, Silvia; Rossi, Alessandra; Esposito, Raffaella. - In: FILOSOFIA MORALE. - ISSN 2785-5457. - 1:7(2025).
14. Rossi, S., D'Alterio, P. (2017). Gaze Behavioral Adaptation Towards Group Members for Providing Effective Recommendations. In: Kheddar, A., et al. *Social Robotics. ICSR 2017. Lecture Notes in Computer Science()*, vol 10652. Springer, Cham. [https://doi.org/10.1007/978-3-319-70022-9\\_23](https://doi.org/10.1007/978-3-319-70022-9_23)

15. Cucciniello, I., Sangiovanni, S., Maggi, G. et al. Mind Perception in HRI: Exploring Users' Attribution of Mental and Emotional States to Robots with Different Behavioural Styles. *Int J of Soc Robotics* 15, 867–877 (2023). <https://doi.org/10.1007/s12369-023-00989-z>
16. Sharkey, A., Sharkey, N. Granny and the robots: ethical issues in robot care for the elderly. *Ethics Inf Technol* 14, 27–40 (2012). <https://doi.org/10.1007/s10676-010-9234-6>
17. Cameron, D. et al. (2017). Children's Age Influences Their Use of Biological and Mechanical Questions Towards a Humanoid. In: Gao, Y., Fallah, S., Jin, Y., Lekakou, C. (eds) *Towards Autonomous Robotic Systems. TAROS 2017. Lecture Notes in Computer Science()*, vol 10454. Springer, Cham. [https://doi.org/10.1007/978-3-319-64107-2\\_23](https://doi.org/10.1007/978-3-319-64107-2_23)
18. Rossi, S., Di Napoli, C., Garramone, F. et al. Personality-Based Adaptation of Robot Behaviour: Acceptability Results on Individuals with Cognitive Impairments. *Int J of Soc Robotics* 16, 211–226 (2024). <https://doi.org/10.1007/s12369-023-01074-1>
19. Raffaella Esposito, Alessandra Rossi, and Silvia Rossi. 2025. Deception in HRI and Its Implications: A Systematic Review. *J. Hum.-Robot Interact.* 14, 3, Article 47 (June 2025), 26 pages. <https://doi.org/10.1145/3721297>
20. Bertolini, A., Arian, S. (2022). Automated Care-Taking and the Constitutional Rights of the Patient in an Aging Population. In: Custers, B., Fosch-Villaronga, E. (eds) *Law and Artificial Intelligence. Information Technology and Law Series*, vol 35. T.M.C. Asser Press, The Hague. [https://doi.org/10.1007/978-94-6265-523-2\\_16](https://doi.org/10.1007/978-94-6265-523-2_16)
21. R. Esposito, A. Rossi, M. Ponticorvo and S. Rossi, "RoboLeaks: Non-strategic Cues for Leaking Deception in Social Robots," 2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Melbourne, Australia, 2025, pp. 1111-1120, doi: 10.1109/HRI61500.2025.10974221.

## Agency e tecnologie cognitive. Uno sguardo al rapporto tra IA e lavoro – di Sebastiano Bavetta, Dipartimento SEAS, Università degli Studi di Palermo

*Il dibattito sul rapporto tra IA e lavoro non riguarda solo la tecnologia, ma anche le modalità di utilizzo. Agency e interazioni determinano se l'IA sia complementare o sostitutiva. Un nuovo indice fornisce dati concreti sul fenomeno*

Quanto lavoro umano è sostituito dall'intelligenza artificiale? Ecco una domanda importante priva di risposta non per mancanza di opinioni, ma di misure. Infatti, per misurare la sostituzione uomo-macchina dovremmo sapere, per ogni contesto lavorativo, *chi* stia eseguendo il compito (essere umano, IA, o entrambi), *quanto* la tecnologia intervenga sui diversi compiti del lavoro umano, *come* modifichi la produttività, la qualità e la centralità del lavoro umano. Informazioni sul *chi*, *quanto* e *come* sono necessarie perché delineano l'impegno cognitivo dell'utente nell'utilizzazione dell'intelligenza artificiale per svolgere un dato compito.

Permettetemi un esempio. Un'insegnante di liceo che ha poco tempo si rivolge ad un modello linguistico per preparare una lezione sul canto di Ulisse nella Divina Commedia. Con un prompt secco chiede al modello di preparargli alcune slide di una presentazione per la sua classe di liceo, includendovi un'introduzione, un'analisi del testo e del suo significato, una contestualizzazione storica e una conclusione.

In questo caso il *chi* è l'insegnante, il *quanto* – l'occupazione orizzontale dei compiti del lavoro da parte della macchina – è significativo poiché praticamente ogni compito necessario per preparare la lezione è realizzato dall'IA, e il *come* – cioè l'uso creativo o sostitutivo della tecnologia – è rivelatore di un basso livello di *agency* da parte dell'insegnante poiché l'output è ricevuto per lo più passivamente.

**Agency:** In filosofia dell'azione e nelle scienze cognitive, il concetto di *agency* indica la capacità di un soggetto di agire intenzionalmente, di darsi fini, di intervenire sul mondo in modo riflessivo e responsabile.

Nel contesto dell'intelligenza artificiale, avere *agency* significa non limitarsi a ricevere passivamente l'output del modello, ma orientare, guidare e valutare la collaborazione uomo-macchina.

### Box 1: Il concetto di *agency*.

L'insegnante avrebbe potuto interagire diversamente con il modello. Avrebbe potuto approcciare il prompting con una sua tesi, per esempio, presentare Ulisse come archetipo dell'uomo moderno che sfida l'ignoto, e far dialogare Dante con le grandi domande dell'era digitale. Avrebbe quindi potuto esplorare, con un prompting serrato con il modello, le connessioni tra l'Ulisse dantesco e il tema dell'innovazione tecnologica circoscrivendo, dopo ciascuna risposta, l'oggetto del suo ragionamento

con richieste volte a plasmare e raffinare l'output ottenuto per renderlo coerente con il messaggio desiderato.

Il *quanto*, e il *come* del rapporto tra lavoro umano e IA sarebbero stati diversi. L'IA avrebbe arricchito, stimolato, fornito materia grezza, senza occupare tutti i compiti del lavoro poiché l'insegnante avrebbe tenuto fermamente in mano il timone facendo della tecnologia un uso creativo e incrementale. A dispetto della similarità delle circostanze, i due casi descrivono interazioni tra il lavoro umano e l'IA che si portano dietro conseguenze diverse, riassumibili, in queste pagine, in una variabile che distingue i due casi: il livello di *agency* profusa dall'insegnante.

Più *agency* l'utente riversi nell'utilizzazione dell'IA più la tecnologia è incrementale invece che sostitutiva perché la capacità di sostituire il lavoro non è una caratteristica intrinseca della tecnologia. L'*agency* è quindi il discriminante. Inoltre, a dispetto dei timori sollevati dalle caratteristiche tecnologiche dell'intelligenza artificiale, oggi essa è sostitutiva con il consenso degli esseri umani. Una nuova base informativa, il *Claude Economic Index*, consente di difendere entrambe le affermazioni alla luce dell'effettiva utilizzazione che chi lavora fa dell'IA.

Capacità incrementale (*augmentation*) e automazione sono due modi distinti di usare l'intelligenza artificiale nel lavoro.

- ◆ Capacità incrementale: l'IA supporta il lavoro umano in un compito. Fornisce stimoli, proposte, bozze, che il lavoro umano poi elabora, adatta, migliora, mantenendo il controllo e la direzione. Osserviamo capacità incrementale soprattutto nello svolgimento di compiti creativi, analitici, strategici

## Box 2: Capacità incrementale e automazione.

### Misure di impatto sul lavoro

Per diventare dirimente tra capacità incrementale e sostituzione del lavoro, l'*agency* deve superare alcune sfide. La prima è quantitativa: trovare una buona misura. La partenza è già in salita. Gli strumenti che la scienza economica usa per misurare l'impatto dell'IA sul lavoro non catturano la dimensione cognitiva del lavoro umano nell'utilizzazione dell'intelligenza artificiale. Essi ritengono che il discriminante tra capacità incrementale e sostituzione dipenda dalle caratteristiche della tecnologia e non osservano il comportamento umano. Vediamo brevemente questi strumenti dell'economia e i loro limiti.

**2.1. Elasticità di sostituzione.** Il primo strumento è l'elasticità di sostituzione tra capitale e lavoro [1]. Essa misura quanto facilmente un'impresa (o un'economia) possa rimpiazzare il lavoro umano con capitale (macchine, infrastrutture, tecnologie), mantenendo lo stesso livello di output. In generale, se la tecnologia migliora, sarà più capace di sostituire il lavoro, con effetti sulla distribuzione dei benefici

derivanti dal progresso tecnologico, sulla disuguaglianza dei redditi e sulla capacità dell’impresa (o dell’economia) di portare a compimento compiti cognitivi complessi.

Il grado di sostituibilità (indicato con  $\sigma$ ) misura quanto cambi il rapporto tra capitale e lavoro (cioè quanta tecnologia sia usata al posto delle persone) quando cambi il prezzo relativo dei due fattori, cioè il salario rispetto al costo del capitale. In altre parole, se  $\sigma$  fosse maggiore di 1 e il salario crescesse, l’impiego di capitale aumenterebbe di più della diminuzione dell’utilizzazione del lavoro, indicando chiaramente un effetto di sostituzione. Se invece  $\sigma$  fosse minore di 1 e il salario crescesse, l’impiego di capitale aumenterebbe meno della diminuzione nella utilizzazione del lavoro, per cui l’effetto di sostituzione sarebbe modesto.

Una prima risposta alla domanda sugli effetti dell’IA sul lavoro è quindi calcolare  $\sigma$ . Senonché, l’intelligenza artificiale non è un bulldozer digitale che sostituisce forza lavoro indistinta. È, invece, una *general-purpose technology* [2] [3] che può sostituire, affiancare, trasformare il lavoro umano in molti modi, secondo caratteristiche che dipendono dalle specifiche circostanze dei compiti da sostituire e dalle attitudini degli utenti. La logica capitale/lavoro è troppo grezza per catturare queste sfumature.

**2.2. Lavoro e compiti.** Dati i limiti dell’elasticità di sostituzione, gli economisti hanno proposto un altro approccio nel quale il lavoro non è un unico fattore indistinto ma un’aggregazione di *compiti* (*task*) che, tutti insieme, lo caratterizzano. Per esempio, un medico diagnostica, prescrive farmaci, ma anche ascolta i pazienti, ne osserva le caratteristiche psicosomatiche e li rassicura. Compiti diversi reagiscono diversamente in termini di sostituibilità all’evoluzione tecnologica per cui il capitale (inclusa l’IA) può sostituire alcuni task ed essere complementare ad altri, anche all’interno dello stesso lavoro [4] [5] [6].

In generale, i compiti più facilmente automatizzabili sono quelli ‘routinari’, vale a dire, ripetitivi, codificabili e descritti da regole ben definite. Questi compiti sono presenti sia nei lavori manuali (assemblaggio in fabbrica), sia in quelli cognitivi (registrazione contabile) e sono più facilmente sostituibili da computer, robot e, ora, modelli di IA. Per contro, i compiti non routinari (pensate al problem solving creativo, all’empatia interpersonale o alla gestione dell’incertezza), sono più resistenti all’automazione e ancorati all’intervento umano qualificato.

Sebbene i modelli task-based abbiano permesso una maggiore precisione nella valutazione del rischio di automazione per diversi tipi di lavoro e lo sviluppo di politiche di formazione mirate, essi assumono che la natura dei task sia fissa. In realtà, l’IA può trasformare i compiti, non solo sostituirli, per cui l’ipotesi di invarianza dei task è difficile da accettare. Soprattutto, dal punto di vista cognitivo, le misure che riguardano i task non osservano l’uso reale della tecnologia e trascurano l’intenzionalità, cioè come il lavoratore scelga di avvalersi dell’IA.

**2.3. AI Exposure Score.** Il terzo strumento di misurazione è recente [7]. Sulla spinta dell’impatto dei modelli di intelligenza artificiale generativa gli economisti hanno cercato strumenti per stimare quali lavori siano più ‘a rischio’ o, per meglio dire, più esposti all’uso dell’intelligenza artificiale.

Diversamente dagli altri due strumenti, in questo approccio non si misura l’effettiva sostituzione ma il suo rischio, attraverso l’*AI Exposure Score*. Attraverso un sofisticato metodo che combina la descrizione delle attività lavorative, delle capacità dell’IA generativa e il matching semantico tra IA e

lavoro umano, lo Score assegna un valore numerico ad ogni occupazione. Il numero riflette il grado di sovrapposizione tra ciò che il lavoro richiede e ciò che l'intelligenza artificiale può fare.

L'AI *Exposure Score* è un'interessante evoluzione nella mappa degli strumenti di misurazione della natura dinamica, adattiva e interattiva dell'intelligenza artificiale nei processi lavorativi, affetta però da un importante limite: lavorando nello spazio del potenziale, non osserva ciò che realmente accade tra IA e lavoro. Ma ciò che accade è la base informativa per una misura che incorpori gli aspetti cognitivi della relazione con l'IA, per cui la prima sfida per l'*agency* non è ancora vinta.

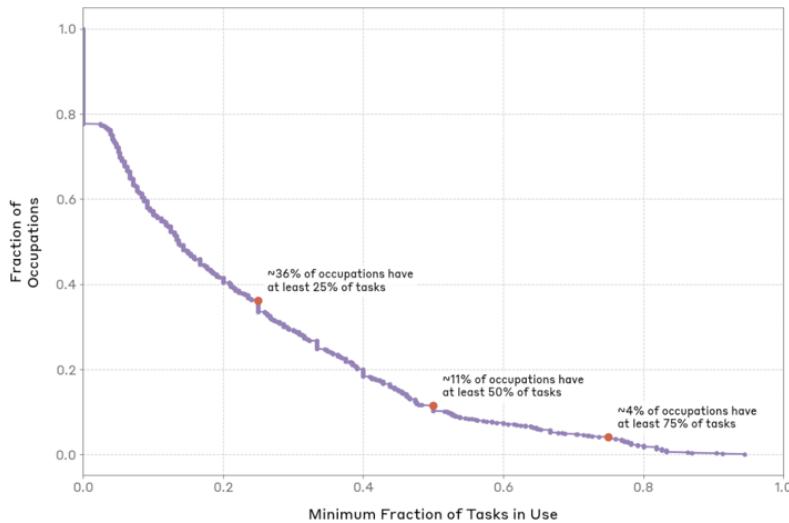
### Misurare l'uso rivelato

**3.1. Il Claude Economic Index.** Un recente articolo, scritto da un team di ricercatori di Anthropic e pubblicato a febbraio 2025 negli ArXiv [8], offre una soluzione all'*agency* per superare la sfida della misura: il *Claude Economic Index*. Combinando milioni di conversazioni anonimizzate tra utenti e il modello Claude, catturate tramite la piattaforma Clio, con i compiti e le occupazioni contenute nel database O\*NET del Dipartimento del Lavoro statunitense, l'indice genera una fotografia dinamica dell'uso dell'intelligenza artificiale: in quali professioni entra di più (*chi*)? In quali compiti (*quanto*)? E soprattutto: con quali modalità (*come*) il lavoratore approccia questa nuova tecnologia, per farsi potenziare o sostituire?

L'innovazione del *Claude Economic Index* è, anzitutto, metodologica. L'obiettivo è puntato non su ciò che accadrà ma su ciò che di fatto accade quando nello svolgere il proprio lavoro si accoglie l'IA. Detto in altri termini, l'unità informativa è l'uso rivelato da parte degli utenti, combinato con la mappa dei lavori e dei compiti nell'economia americana. Emerge così dai dati una nuova prospettiva sull'utilizzazione dell'IA leggibile in chiave di *agency*.

**3.2. Chi usa l'IA?** Anzitutto, *chi* usa l'IA? Nella collaborazione con l'intelligenza artificiale non tutti i lavori sono uguali. L'immenso dataset delle interazioni con Claude rivela che solo il 4% delle occupazioni utilizza l'IA in almeno il 75% dei compiti che compongono ciascuna occupazione, mentre circa il 36% delle occupazioni utilizza l'IA in almeno il 25% dei propri task. Inoltre, la scelta di utilizzazione dipende fortemente dalle caratteristiche tecnologiche dell'intelligenza artificiale poiché queste ultime determinano la probabilità di offrire valore.

Infatti, le professioni che mostrano il grado più alto di interazione con l'IA sono quelle in cui i compiti principali sono verbali, concettuali e strutturati in linguaggio naturale, ovvero digitalizzati o già mediati da interfacce testuali. Il *chi* è quindi un insegnante, un analista o consulente, un operatore del marketing o di contenuti digitali, uno sviluppatore o data scientist.



**Figura 1:** Utilizzo dell'IA nelle diverse professioni. Distribuzione cumulativa che mostra la frazione di professioni (asse y) che utilizza l'IA in almeno una determinata frazione dei propri compiti (asse x).

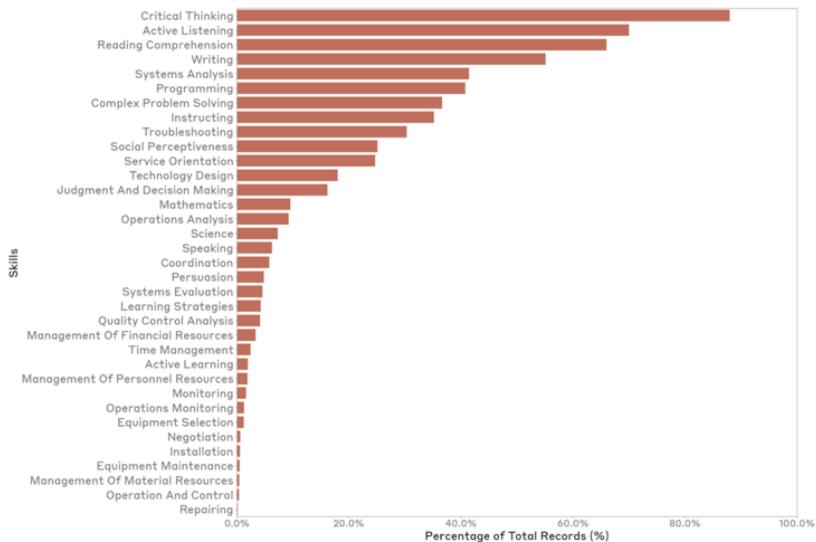
Fonte: Anthropic Claude Economic Index [8].

La Figura 1 descrive la distribuzione cumulativa delle professioni che usano l'intelligenza artificiale per lo svolgimento di almeno alcuni compiti. In alto a sinistra troviamo le professioni (un po' più del 20%, in America) che non usano l'IA. Scendendo verso destra le professioni che, via via, fanno dell'IA un uso più intenso, vale a dire che la utilizzano nell'espletamento di un numero elevato di compiti.

Che cosa ci dice questo grafico? Che l'intelligenza artificiale non sta ancora colonizzando in massa le professioni. Piuttosto, entra in punta di piedi, occupando spazi ben definiti all'interno dei mestieri – ad esempio nei compiti legati alla scrittura o all'analisi – senza rimpiazzare tutto il flusso lavorativo. Si estende, per così dire, a bassa intensità ma ad ampio spettro. Ancora, si osserva che oggi l'IA lavora per compiti, non per interi mestieri. La sostituzione completa dell'essere umano è un fenomeno marginale mentre la collaborazione selettiva è la norma.

**3.3. Quanto si usa l'IA?** Il Claude Economic Index ci dice anche quanto si usi l'IA, cioè la distribuzione delle abilità (e quindi dei compiti) che dominano le conversazioni tra Claude e i suoi utenti.

La distribuzione delle abilità lavorative che emergono più frequentemente nelle conversazioni tra esseri umani e il modello Claude, descritta nella Figura 2, è un indicatore di quali skill gli utenti sollecitino nei compiti quotidiani. La figura rivela una geografia precisa delle complementarietà uomo-macchina. I picchi più alti si osservano in competenze legate al pensiero critico, alla programmazione o alla comprensione dei testi. Negli ambiti che alloggiano nella parte superiore della figura gli utenti attribuiscono alla conversazione con l'IA maggiore utilità e tendono ad usarla estensivamente.



**Figura 2:** Distribuzione delle competenze professionali nelle conversazioni con Claude.

Fonte: Anthropic Claude Economic Index [8].

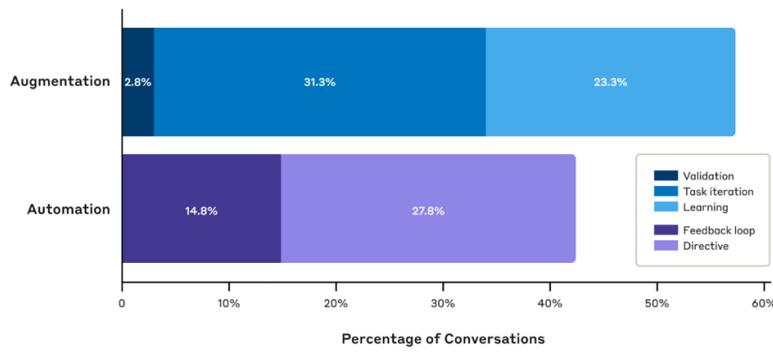
Dai dati emerge che alcune competenze sono sistematicamente marginali nelle conversazioni analizzate tra utenti umani e Claude. Esse si concentrano in due grandi famiglie. La prima include le competenze manuali e tecnico-operative che richiedono interazione fisica diretta con il mondo materiale, spesso in contesti ad alta variabilità. Queste attività sono scarsamente rappresentate perché l'IA generativa opera esclusivamente nel dominio simbolico e linguistico mentre queste competenze richiedono capacità sensoriali, manipolative e contestuali difficilmente codificabili e simulabili nel linguaggio naturale.

La seconda famiglia comprende le competenze interpersonali ad alta densità relazionale che richiedono abilità sociali e comunicative ancora inaccessibili per l'IA. Sono competenze negoziali, di percezione sociale, di influenza nelle relazioni sociali, basate su empatia, intenzionalità e lettura del contesto umano che si manifestano in ambienti informali che le interfaccia digitali non possono catturare.

L'assenza dai piani alti della figura non significa che queste competenze siano al sicuro per sempre, ma che, finché l'interfaccia uomo-macchina resterà linguistica e digitale, ci sarà una fascia significativa di attività umane che resterà non sostituibile, o perlomeno non automatizzabile con efficienza.

**3.4. Come si usa l'IA?** Nel mondo dell'intelligenza artificiale conta anche *come* le persone scelgano di usarla. La Figura 3 propone una mappa dei pattern di comportamento degli utenti in relazione ai modelli linguistici generativi, distinguendo tra automazione e arricchimento funzionale.

Il contributo chiave della figura non è contabile, ma è l'introduzione di una tassonomia dei modi di usare l'IA, organizzata su due poli, *augmentation* (capacità incrementale) e *automation* (sostituzione del lavoro).



**Figura 3.** Distribuzione di comportamenti di automatizzazione (43%) in cui gli utenti delegano compiti all'IA e comportamenti di incremento funzionale (57%) in cui gli utenti collaborano attivamente con l'IA.

Fonte: Anthropic Claude Economic Index [8].

Secondo i dati, il 57% delle interazioni analizzate è di tipo incrementale, mentre il 43% mostra pattern di automatizzazione. Nella maggior parte dei casi gli utenti non si limitano a delegare all'IA, ma la coinvolgono come partner nel processo cognitivo. Tuttavia, la distribuzione varia significativamente tra occupazioni. Alcuni lavori favoriscono l'automazione (per esempio, data entry, produzione testi standardizzati), altri l'incremento funzionale (per esempio, insegnamento, strategia, design). La variabilità riflette la natura del task, ma anche la cultura e il modo di operare dell'ambiente professionale.

Per chiarezza interpretativa, la Tabella 1 descrive analiticamente i comportamenti di automatizzazione e di incremento funzionale introdotti dalla Figura 3.

Comportamenti di automatizzazione	Comportamenti incrementali
<i>Directive</i> : il compito viene dato all'IA senza ulteriore intervento umano	<i>Task Iteration</i> : l'utente dialoga, affina, riscrive
<i>Feedback Loop</i> : l'IA risponde, l'utente approva o chiede piccole revisioni	<i>Learning</i> : l'utente usa l'IA per apprendere o testare concetti
	<i>Validation</i> : l'utente confronta l'output dell'IA con le proprie intuizioni

**Tabella 1.** Tassonomia dei comportamenti dell'utente.

Fonte: Anthropic Claude Economic Index [8].

La misura restituita dalla Figura 3 è essenziale affinché l'*agency* possa superare la prima sfida, per tre motivi. Anzitutto la rende visibile: chi esplora, itera, verifica, esercita controllo e *agency*; chi delega, abdica, copia-incolla, limita l'interazione e permette alla macchina di sostituire il lavoro umano. La seconda ragione è che rende negoziabile l'automazione nel rapporto con la macchina. Mostra che

l'automazione non è un destino imposto dalla macchina o l'ineluttabile conseguenza delle sue caratteristiche tecnologiche, ma una scelta di design o di uso. Ciascun utente sceglie se automatizzare o utilizzare la tecnologia a fini incrementali. Infine, aiuta a disegnare policy e formazione perché contribuisce a comprendere quali pattern nella relazione tra utente e IA prevalgano in un dato settore.

## L'IA è una tecnologia aperta

La seconda sfida per l'*agency* è concettuale. I modelli economici sull'impatto dell'IA sul lavoro assumono che la tecnologia, una volta disponibile, abbia un potere intrinseco di sostituzione del lavoro umano, in particolare laddove essa sia più efficiente o qualitativamente preferibile rispetto all'uomo. Ancora una volta nei modelli economici il *come* dipende da elementi oggettivi delle caratteristiche della tecnologia che hanno impatto sulla produttività e sulla qualità del prodotto e non dalle scelte dell'utente. L'esclusione dell'*agency* dalla valutazione del *come* regge però se vale il presupposto che le tecnologie siano chiuse, cioè progettate per uno scopo univoco e indipendente dall'uso umano.

**4.1. Giudizi ‘pre-specificati’ e ‘complemento’.** Ma è davvero così? Con l'introduzione dell'IA il presupposto è discusso intensamente nella scienza economica, in risposta al timore che le peculiari caratteristiche tecnologiche dell'intelligenza artificiale accelerino inesorabilmente la sostituzione del lavoro. In un recente articolo, Agrawal, Gans e Goldfarb [9] costruiscono un modello formale di interazione con una tecnologia cognitivamente abilitante in cui individui motivati dal desiderio di migliorare la realizzazione dei propri task fanno uso della tecnologia per raggiungere il proprio obiettivo (osservate che questa è affermazione equivale a dire che aumenta la produttività).

Il modello è interessante per diverse ragioni. Anzitutto, quando si adotta una tecnologia cognitiva abilitante, gli agenti riducono il livello di impegno nello svolgimento dei task poiché la tecnologia aumenta la probabilità di raggiungere buoni risultati. Quindi, un effetto di sostituzione direttamente legato alle caratteristiche abilitanti della tecnologia è parte dei risultati del modello perché la sostituzione consente di raggiungere un maggior beneficio netto anche a fronte di un minore impegno. In altri termini, il miglioramento della produttività è legato in qualche modo alla sostituzione.

D'altra parte, l'introduzione di una tecnologia cognitiva abilitante non può portare solo a sostituzione (automazione, dicono gli autori). La ragione è che l'automazione richiede un *pre-specified judgement*, cioè una decisione o una valutazione già determinata *ex ante* nel processo operativo. Per esempio, se un algoritmo deve applicare una formula già nota a un dataset (pensate al calcolo dell'IVA), non serve alcun giudizio da parte dell'utente perché l'output perseguito è meccanico. Gli autori dicono che il giudizio è pre-specificato.

Ma una tecnologia cognitiva abilitante non si applica solo a circostanze in cui il giudizio sia pre-specificato. L'insegnante che prepari la lezione sul Canto di Ulisse non può utilizzare dei giudizi pre-specificati ma userà comunque un giudizio che, nel suo caso, è un complemento della tecnologia cognitiva. Di conseguenza, quando l'IA si trova di fronte a giudizi dai confini imprecisi che servono da complemento per realizzare il suo prodotto, allora l'automazione completa diventa meno probabile.

Inoltre, man mano che la tecnologia cognitiva diventa più potente, la probabilità di automazione diminuisce ulteriormente perché cresce il valore del giudizio complementare offerto dal lavoro umano poiché i compiti diventano più sofisticati e il margine di errore meno tollerabile. Nel Box 3 riporto un altro esempio del ruolo della complementarietà del giudizio.

Un ospedale utilizza un modello di intelligenza artificiale per analizzare immagini di risonanza magnetica. In uno scenario incrementale in cui il giudizio non è pre-specificato, il medico riceve un suggerimento dall'IA, ma poi valuta la proposta alla luce della storia clinica del paziente, di altri esami diagnostici e del colloquio con il malato. Qui, l'IA è uno strumento che amplia la capacità del medico di vedere più alternative e di prendere decisioni più informate. Il giudizio umano resta centrale: l'automazione non si realizza.

In uno scenario opposto, l'ospedale delega interamente la decisione all'IA. Se il punteggio di rischio supera una certa soglia, il paziente viene avviato all'intervento chirurgico. Il giudizio è stato pre-specificato, e l'automazione si realizza completamente. Ma se l'IA ha sbagliato – per dei bias nei dati o per la presenza di un pattern anomalo – l'errore non è intercettabile. Il sistema è fragile perché è allo stesso tempo troppo potente per essere ignorato e troppo automatico per essere controllato.

**Box 3:** Un esempio di giudizio non pre-specificato e la sua relazione con l'*agency*.

**4.2. Il ruolo dell'*agency* nel giudizio complemento.** Sebbene Agrawal, Gans e Goldfarb attribuiscano alla relazione tra giudizio sulle circostanze e strumento cognitivo il ruolo di discriminante tra automazione e capacità incrementale, il ruolo dell'*agency* è apparente. Infatti, è proprio l'*agency* che sostiene la complessità e la complementarietà del giudizio umano poiché consente di interpretare il contesto, porre domande nuove, valutare alternative e assumersi la responsabilità della decisione. Senza *agency* il giudizio tornerebbe ad essere pre-specificato e la tecnologia ad essere meramente sostitutiva.

## Conclusione

A dispetto della vasta letteratura distopica che sottolinea i rischi cui il futuro del lavoro sarebbe sottoposto dall'espansione dell'IA, il Claude Economic Index offre un momento di sobrietà ben fondato sui dati e su una metodologia di analisi – quella dell'utilizzazione rivelata dell'IA – innovativa e informativa. Il peculiare punto di osservazione dell'indice permette anche di mettere alla prova il ruolo dell'*agency* come discriminante tra sostituzione e incremento del lavoro umano.

Il risultato che i dati dell'indice permettono di raggiungere, peraltro, non è esclusivamente dipendente dai numeri che esso mette in mostra. L'ipotesi che l'*agency* sia un'informazione dirimente per determinare se l'IA agisca come forza sostitutiva o capacità incrementale trova infatti conferma (indiretta) in altre evidenze empiriche.

Brevemente, eccone una, interessante perché lega l'*agency* alla percezione che gli utenti dell'IA hanno di quanto la tecnologia potrebbe migliorare la propria produttività. Peng, Kalliamvakou, Cihon, e Demirer [10] hanno condotto un esperimento controllato con GitHub Copilot. L'esperimento consiste nel chiedere ad alcuni sviluppatori di implementare un server http in JavaScript il più

velocemente possibile. Un gruppo (il gruppo di trattamento) ha avuto la possibilità, come il vincitore di una Mistery Box di Masterchef, di ricevere istruzioni su come farlo in Copilot. All’altro gruppo, il gruppo di controllo, è stata data solo la possibilità di acquisire informazioni attraverso una demo. Il primo gruppo ha una maggiore percezione del miglioramento che può realizzare nel completare il compito, maggiore persino del miglioramento che di fatto realizza.

L’effetto cognitivo indotto dall’utilizzazione dell’IA va oltre ciò che gli utenti riescono a registrare consapevolmente [11]. Sebbene non abbiano informazioni sul livello di agency di ciascuno dei componenti dei due gruppi, la sovrastima fatta dal gruppo sottoposto al trattamento rivela che la maggiore produttività non dipenda solo dal prompt iniziale, bensì da un processo d’interazione più profondo che chiama a raccolta le abilità cognitive delle persone.

Anche se non adeguatamente esplorata nelle analisi sul futuro del lavoro, né empiricamente osservabile, l’agency è un moltiplicatore di efficacia capace di sfruttare al meglio i benefici dell’IA. Ed oggi, anche la diga migliore all’avanzamento della sostituzione.

## Riferimenti bibliografici

[1] Daron Acemoglu, Pascual Restrepo (2020). “Robots and Jobs: Evidence from US Labor Markets.” *Journal of Political Economy*, Volume 128, Pages 2188-2244.

[2] Timothy Bresnahan, Manuel Trajtenberg (1995). “General Purpose Technologies ‘Engines of growth?’”, *Journal of Econometrics*, 65, Pages 83-108.

[3] Timothy Bresnahan (2010). “Chapter 18 - General Purpose Technologies”, in Bronwyn H. Hall, Nathan Rosenberg (a cura di), *Handbook of the Economics of Innovation*, North-Holland, Volume 2, Pages 761-791.

[4] David H. Autor, Frank Levy, Richard J. Murnane, (2003). The Skill Content of Recent Technological Change: An Empirical Exploration, *The Quarterly Journal of Economics*, Volume 118, Pages 1279–1333, <https://doi.org/10.1162/003355303322552801>

[5] Daron Acemoglu, David Autor (2011). “Skills, Tasks and Technologies. Implications for Employment and Earnings”. *Handbook of Labor Economics*, Volume 4, Part B, Pages 1043-1171, Elsevier.

[6] Daron Acemoglu, Pascual Restrepo, (2019). “Artificial Intelligence, Automation and Work.” In Ajay Agrawal, Joshua Gans, Avi Goldfarb (a cura di), *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press, Pages 197-236.

- [7] Erik Brynjolfsson, Danielle Li, Lindsey R. Raymond (2023). “Generative AI at Work.” *NBER Working Paper* No. 31161.
- [8] Kunal Handa, Alex Tamkin, Miles McCain, Saffron Huang, Esin Durmus, Sarah Heck, Jared Mueller, Jerry Hong, Stuart Ritchie, Tim Belonax, Kevin K. Troy, Dario Amodei, Jared Kaplan, Jack Clark, Deep Ganguli (2025). “Which Economic Tasks are Performed with AI? Evidence from Millions of Claude Conversations”. *ArXiv* 2503. 04761v1.
- [9] Ajay K. Agrawal, Joshua S. Gans, Avi Goldfarb (2025). “The Economics of Bicycles for the Mind.” *NBER Working Paper* No. 34034.
- [10] Sida Peng, Eirini Kalliamvakou, Peter Cihon, Mert Demirer (2023). “The Impact of AI on Developer Productivity: Evidence from GitHub Copilot.” *arXiv* 2302.06590v1.
- [11] McKinsey & Company (2025). *Superagency in the workplace: Empowering people to unlock AI’s full potential at work.*

# **Conversazioni artificiali in psicologia: strumenti, illusioni e nuove domande cliniche – di Raffaella I. Rumiati, SISSA Trieste**

*L'uso crescente dell'intelligenza artificiale in psicologia apre opportunità di accesso e supporto, ma solleva interrogativi clinici, etici ed epistemologici. App, chatbot e piattaforme ibride richiedono valutazioni critiche per distinguere simulazioni digitali da autentiche pratiche terapeutiche*

## **Abstract**

L'articolo prende in esame l'uso crescente dell'intelligenza artificiale in ambito psicologico, con particolare attenzione alle pratiche di *self-help* e ai sistemi automatizzati di supporto emotivo (come chatbot conversazionali, app per la regolazione affettiva, meditazione e *journalling* digitale). Dopo aver ricostruito l'attuale ecosistema tecnologico, discute criticamente i presupposti teorici e le implicazioni etiche, cliniche ed epistemologiche, suggerendo infine possibili correttivi per un uso consapevole e rispettoso dei principi della cura psicologica.

## **Il boom della salute mentale digitale**

Negli ultimi anni, l'aumento dei livelli di ansia, depressione e stress cronico, aggravati dalla pandemia, ha fatto emergere una crescente domanda di supporto psicologico. È così che ha preso forma un panorama sempre più articolato di strumenti digitali dedicati al benessere mentale: app di mindfulness, chatbot empatici, piattaforme di telepsicologia, simulatori di conversazione affettiva. Queste soluzioni promettono un accesso immediato e flessibile, meno condizionato dallo stigma che spesso accompagna la richiesta di aiuto psicologico.

Tuttavia, il successo di queste soluzioni non si spiega soltanto con il loro appeal tecnologico o con la retorica dell'autonomia individuale. Molti utenti vi ricorrono per ragioni pragmatiche: la psicoterapia tradizionale comporta dei costi, richiede tempo e spostamenti ed è spesso difficile da conciliare con i ritmi della vita quotidiana. Inoltre, l'accesso ai servizi psicologici continua a essere disomogeneo sul territorio e meno garantito per le fasce sociali più svantaggiate. Iniziative come il bonus psicologico vanno nella direzione di una maggiore equità, ma i fondi disponibili e le procedure di accesso non sono sempre sufficienti a coprire il reale fabbisogno della popolazione.

In questo contesto, la disponibilità immediata e la fruizione in assenza di un rapporto diretto con un professionista offerte da app e piattaforme digitali rappresentano una risposta concreta a ostacoli strutturali reali, contribuendo almeno in parte a colmare il divario tra domanda e offerta di supporto psicologico.

Tuttavia, queste nuove procedure hanno cominciato a produrre un crescente disorientamento, sia da parte degli utenti sia tra i professionisti. Dobbiamo quindi interrogarci su quali siano le differenze tra un trattamento psicologico condotto da un professionista e un'esperienza digitale di supporto emotivo. Quali sono i limiti di questo approccio? E quali implicazioni comporta l'affidare a un algoritmo una funzione che, per sua natura, è relazionale, contestuale e umana?

In questo articolo mi propongo di analizzare l'attuale offerta digitale per la salute mentale, esaminando le tecnologie oggi disponibili, i modelli concettuali di riferimento e le principali implicazioni etiche, epistemologiche e cliniche. Pur trattandosi di un fenomeno ormai ben avviato, è

necessario interrogarsi su quali correttivi adottare per garantirne un uso consapevole e responsabile, in linea con i principi fondamentali della cura psicologica. Ma cosa offre oggi il mercato?

### **App di self-help e mindfulness**

Tra gli strumenti più diffusi nell'ecosistema digitale si trovano le app di self-help e mindfulness, pensate per promuovere una forma di benessere psicologico ‘fai-da-te’ le più diffuse sono *Headspace* e *Calm*.

*Headspace*, fondata nel 2010 da Andy Puddicombe (un ex monaco buddista) e Richard Pierson, è nata nel Regno Unito e si è poi affermata a livello globale grazie a un'efficace campagna mediatica, che include anche una partnership con Netflix. L'app propone percorsi audio di meditazione guidata, esercizi di respirazione, tecniche di rilassamento e contenuti per migliorare la concentrazione e la qualità del sonno.

*Calm*, invece, è sorta nel 2012 sul suolo americano, a San Francisco, ed è focalizzata sul rilassamento e sul sonno, e include una vasta raccolta di “sleep stories”, brani musicali, suoni naturali e meditazioni ispirate alla mindfulness. Il suo design, curato e immersivo, è pensato per evocare benessere e comfort sensoriale, e si colloca più nel mercato del wellness (*lifestyle philosophy*) che in un contesto terapeutico vero e proprio.

In conclusione, entrambe le app sono ispirate alla psicologia positiva e si presentano come strumenti di self-help accessibili e user-friendly, con un'estetica curata e un linguaggio rassicurante. Anche se i loro contenuti sono sviluppati o validati da professionisti del settore, tuttavia, mancano prove solide sulla loro efficacia clinica (cfr. Koh et al., 2022; Weisel et al., 2019). Le collaborazioni con istituzioni scientifiche, spesso dichiarate dalle aziende, raramente producono dati accessibili o pubblicazioni peer-reviewed. Le metriche utilizzate si concentrano sull'engagement e sul gradimento, non su indicatori clinicamente rilevanti. Inoltre, l'assenza di follow-up strutturati e di valutazioni longitudinali impedisce di verificare l'impatto reale di queste app sul benessere psicologico degli utenti.

### **Chatbot conversazionali e relazioni simulate**

L'uso dell'intelligenza artificiale nella salute mentale ha radici lontane. Già nel 1966, Joseph Weizenbaum sviluppò *ELIZA*, uno dei primi tentativi di simulare il linguaggio naturale in un contesto terapeutico automatizzato. Il sistema era decisamente rudimentale, ma sufficiente a generare l'illusione di essere compresi. Oggi, i modelli linguistici di nuova generazione (LLM) rendono questa simulazione molto più realistica. Eppure la domanda sollevata da Weizenbaum resta centrale: quanto possiamo davvero fidarci di un'intelligenza che sembra capirci?

Accanto alle app di self-help, si è diffusa una seconda categoria di strumenti digitali basati su interazione linguistica: i chatbot conversazionali. Tra i più noti si trovano *Woebot*, *Wysa* e *Replika*, applicazioni che simulano il dialogo con un interlocutore empatico, offrendo supporto emotivo e suggerimenti comportamentali. L'esperienza dell'utente si costruisce intorno a scambi testuali, e in alcuni casi vocali, che mirano a stabilire una relazione di ascolto percepita come personale.

*Woebot*, sviluppato da un team di psicologi della Stanford University, negli Stati Uniti, è ispirato alla terapia cognitivo-comportamentale (CBT, Cognitive Behavioural Therapy) e guida l'utente in brevi

interazioni volte a riconoscere distorsioni cognitive e a promuovere strategie di autoregolazione. Rispetto a Woebot, Wysa, anch'esso basato sulla CBT, offre inoltre opzioni di supporto umano (a pagamento), oltre a esercizi di mindfulness, diari emotivi e strumenti per il tracciamento dell'umore.

*Replika*, invece, si presenta come un “AI companion” generalista, orientato alla costruzione di una relazione emotiva continuativa, spesso caratterizzato da toni confidenziali o affettivi. A differenza degli altri due, *Replika* si spinge esplicitamente sul terreno della simulazione relazionale e può essere personalizzato anche in senso romantico o sessualizzato.

Pur offrendo un’interfaccia accessibile e una risposta immediata per chi cerca sollievo emotivo, questi strumenti sollevano interrogativi di natura clinica, metodologica ed etica. La letteratura disponibile su Woebot e Wysa riporta effetti positivi su alcuni sintomi depressivi e ansiosi, ma tali risultati si basano prevalentemente su studi a breve termine, condotti su campioni non clinici e con disegni sperimentali metodologicamente fragili (cfr. Fitzpatrick et al., 2017; Siddals et al., 2024; Coghlan et al., 2023).

*Replika*, invece, è stato oggetto di diverse analisi per la natura del legame affettivo che può instaurarsi tra utente e chatbot. Uno studio longitudinale ha mostrato che, in seguito alla rimozione delle funzionalità romantiche da parte della piattaforma, molti utenti hanno reagito con vissuti di perdita e frustrazione, segnalando un’interruzione del senso di continuità identitaria associato alla relazione con l’AI (De Freitas et al., 2024). Questi dati confermano il rischio di dipendenza emotiva e confusione affettiva in interazioni che, pur non essendo relazioni autentiche, possono assumere un ruolo centrale nell’equilibrio psicologico dell’utente.

Accanto ai nomi più noti, si diffondono anche companion informali come *Broken Bear AI*, che promettono conforto emotivo e continuità affettiva senza finalità terapeutiche dichiarate. Anche in questi casi, la relazione simulata ricalca il modello della cura, ma senza contesto clinico o supervisione, rendendo sempre più sfumata la distinzione tra empatia artificiale e supporto reale.

In sintesi, ciò che i chatbot propongono come empatia è una simulazione priva di coscienza, intenzionalità e contesto umano: si limita a riprodurre pattern linguistici appresi, senza alcuna forma di comprensione autentica. I dati generati durante le interazioni con i chatbot, spesso riferiti a emozioni, stati mentali e vissuti personali, sono altamente sensibili. Tuttavia, vengono spesso raccolti, archiviati e utilizzati secondo criteri poco trasparenti. Questa opacità solleva preoccupazioni significative per la privacy, la protezione dei dati e il diritto degli utenti a sapere come vengono trattate le proprie informazioni psicologiche (Luxton, 2022).

### **Soluzioni ibride: tra algoritmi e terapeuti**

Una terza categoria è rappresentata dalle piattaforme ibride – come Youper, Talkspace e BetterHelp – che combinano algoritmi digitali con l’intervento diretto di professionisti della salute mentale. Questi servizi si presentano come alternative flessibili alla psicoterapia tradizionale, promettendo percorsi personalizzati, accesso rapido e costi più contenuti.

*Youper*, ad esempio, che integra un chatbot basato sulla CBT con la possibilità di consulti umani, ha mostrato benefici significativi su ansia e depressione entro le prime due settimane dell’utilizzo (Mehta et al., 2021). *Talkspace* e *BetterHelp* offrono sedute condotte da terapeuti abilitati e, in studi longitudinali su campioni consistenti, hanno riportato miglioramenti clinici, soprattutto per quanto concerne la depressione e l’ansia. Tuttavia, l’efficacia di questo approccio rispetto alla tradizionale terapia in presenza resta ancora da valutare (Mehta et al., 2021; Marcelle et al., 2019).

Sebbene queste piattaforme rispondano a un bisogno reale di accessibilità, in particolare in contesti geografici o sociali dove i servizi psicologici sono carenti, sollevano diverse criticità. La relazione terapeutica, mediata da interfacce digitali e protocolli standardizzati, tende a risultare meno efficace e impersonale. La qualità del trattamento ovviamente dipende in larga parte dalla formazione dei terapeuti, specialmente se inseriti in questi protocolli ibridi.

A questo si aggiungono interrogativi non secondari sulla sicurezza dei dati clinici, la trasparenza degli algoritmi di accoppiamento paziente-terapeuta e il quadro giuridico e deontologico entro cui operano tali piattaforme.

A fronte di queste criticità, sarebbe opportuno prevedere protocolli di valutazione iniziale, in grado di valutare il profilo psicologico e l'eventuale rischio psicopatologico dell'utente prima dell'avvio del percorso. L'assenza di uno screening strutturato compromette la possibilità di formulare un'indicazione clinica adeguata e aumenta il rischio di trattamenti non appropriati, ritardi nell'invio a percorsi terapeutici efficaci o gestione non conforme alla complessità del caso.

Un esempio emblematico è riportato da Jess McAllen su *The Baffler*, che descrive la sua interazione diretta con il chatbot *Broken Bear*. Quando scrive di non riuscire a smettere di controllare il forno, sintomo tipico del disturbo ossessivo-compulsivo, il sistema risponde: "Oh dear, did you leave it on? Maybe you could check it. But do be careful." Una risposta empatica solo in apparenza, che rafforza esattamente il comportamento disfunzionale che una terapia dovrebbe contrastare.

Un ruolo altrettanto delicato è giocato dai tratti narcisistici, spesso sottovalutati nel contesto delle interazioni digitali. Caratterizzati da una forte sensibilità al riconoscimento esterno e da una regolazione affettiva centrata sul rispecchiamento, tali tratti possono trovare nei chatbot una fonte di gratificazione immediata ma illusoria. Le simulazioni empatiche offerte dall'IA, sprovviste di reale alterità, sempre disponibili e rassicuranti, rischiano di consolidare modalità relazionali egocentriche e dipendenti, invece di stimolare una rielaborazione autentica del Sé in rapporto all'altro. Esempi come questo rendono evidenti i limiti clinici di un'intelligenza artificiale non supervisionata.

In sintesi, queste soluzioni ibride rappresentano una nuova forma di *outsourcing* terapeutico, a metà strada tra medicina personalizzata e meccanismi tipici delle piattaforme digitali come efficienza, algoritmi e standardizzazione (Torous & Roberts, 2017). Un modello indubbiamente promettente in termini di flessibilità e inclusività, ma che necessita oggi di maggiori garanzie in termini di trasparenza, regolazione e valutazione indipendente.

### **Conclusioni critiche: accessibilità, rischi relazionali e scenari regolatori**

Gli strumenti digitali per la salute mentale comprendono soluzioni molto eterogenee, ma vengono spesso presentati sotto un'unica etichetta di "supporto psicologico". Il problema non è la tecnologia in sé, bensì la sua capacità di simulare una relazione di cura, attraverso risposte affettive, interfacce conversazionali e protocolli standardizzati che possono indurre coinvolgimento emotivo, attaccamento o dipendenza.

In assenza di un contesto clinico, la distinzione tra supporto percepito e cura effettiva si fa sempre più sfumata. Il pericolo è duplice: da un lato, che questi strumenti sostituiscano impropriamente percorsi terapeutici necessari; dall'altro, che alimentino narrazioni interiori colpevolizzanti o

deresponsabilizzanti, attribuendo all'utente il fallimento di un processo che era privo, sin dall'inizio, di una reale dimensione terapeutica.

Non si tratta di demonizzare queste tecnologie, né tantomeno di idealizzarle. Si tratta piuttosto di sviluppare un approccio critico, regolato e consapevole, capace di valorizzare le potenzialità di accessibilità e supporto, ma anche di riconoscere i limiti strutturali. Strumenti che restano algoritmici, impersonali e privi di supervisione clinica non possono, e non devono, sostituire la specificità della relazione terapeutica.

## Bibliografia

- Coghlan, S., Leins, K., Sheldrick, S., Cheong, M., Gooding, P., & D'Alfonso, S. To chat or bot to chat: Ethical issues with using chatbots in mental health. *Digit Health*. 2023 Jun 22;9:20552076231183542. doi: 10.1177/20552076231183542. PMID: 37377565; PMCID: PMC10291862.
- Fitzpatrick, K.K., Darcy, A., & Vierhile, M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Ment Health* 2017;4(2):e19 doi: [10.2196/mental.7785](https://doi.org/10.2196/mental.7785)
- De Freitas, J., Castelo, N., Uğuralp, A. K., & Oğuz-Uğuralp, Z. (2024). Lessons from an App Update at Replika AI: Identity Discontinuity in Human–AI Relationships. Harvard Business School Working Paper No. 25-018. arXiv:2412.14190
- Luxton, D.D. (2022). Ethical and Legal Considerations in Artificial Intelligence Applications in Mental Healthcare. In D. D. Luxton (Ed.), *Artificial Intelligence in Behavioral and Mental Health Care* (pp. 247–268). Elsevier. DOI: 10.1016/B978-0-12-821133-0.00012-4
- Marcelle, E.T., Nolting, L., Hinshaw, S.P., & Aguilera, A. (2019). Effectiveness of a multimodal digital psychotherapy platform for adult depression: A naturalistic feasibility study. *JMIR mHealth and uHealth*, 7(1), e10948. <https://doi.org/10.2196/10948>
- McAllen, J. The therapist in the machine. *The Baffler*, N.76, November 2024. Trad. it. Il terapeuta artificiale, *Internazionale*, n. 1599, 31 gennaio
- Mehta A, Niles AN, Vargas JH, Marafon T, Couto DD, Gross JJ. Acceptability and Effectiveness of Artificial Intelligence Therapy for Anxiety and Depression (Youper): Longitudinal Observational Study. *J Med Internet Res*. 2021 Jun 22;23(6):e26771. doi: 10.2196/26771.
- Koh, J., Tng, G.Y.Q., & Hartanto, A. Potential and pitfalls of mobile mental health apps in traditional treatment: An umbrella review. *J Pers Med*. 2022 Aug 25;12(9):1376. doi: 10.3390/jpm12091376.
- Siddals, S., Torous, J. & Coxon, A. “It happened to be the perfect thing”: experiences of generative AI chatbots for mental health. *npj Mental Health Res* 3, 48 (2024). <https://doi.org/10.1038/s44184-024-00097-4>
- Torous, J. & Roberts, L.W. (2017). The ethical use of mobile health technology in clinical psychiatry. *Journal of Nervous and Mental Disease*, 205(1), 4–8.  
DOI: 10.1097/NMD.0000000000000596

Weisel, K.K., Fuhrmann, L.M., Berking, M. *et al.* Standalone smartphone apps for mental health—a systematic review and meta-analysis. *npj Digit. Med.* 2019, **2**, 118, <https://doi.org/10.1038/s41746-019-0188-8>

## **Le competenze chiave per l'intelligenza artificiale generativa: verso un approccio *human-centered* al governo della AI – di Giovanni Battista Dagnino, Professore ordinario di Digital Strategy & Artificial Intelligence**

L'intelligenza artificiale generativa (GenAI) ha fatto registrare una crescita esplosiva negli ultimi anni, trasformandosi da semplice curiosità tecnologica a oggetto di impiego quotidiano per centinaia di milioni di persone. Secondo i dati dell'Osservatorio sulla Digital Innovation del Politecnico di Milano, in Italia nel 2024, il mercato della AI ha raggiunto un valore record di 1,2 miliardi di euro con una crescita del 58% rispetto all'anno precedente 2023. Questo importante incremento testimonia non soltanto l'interesse crescente verso le nuove tecnologie *AI-driven*, ma anche la loro progressiva integrazione nella vita quotidiana e nei processi aziendali. Il 28% degli utenti internet in Italia, circa 13 milioni di persone, ha utilizzato almeno un'applicazione di intelligenza artificiale generativa nel maggio 2025. Il protagonista indiscusso di questa rivoluzione digitale è ChatGPT, utilizzato da 11 milioni di italiani, con un profilo utente prevalentemente giovane: il 44,5% ha tra i 15 e i 24 anni e il 51,9% è di sesso femminile. Ancora più impressionante l'intensità di utilizzo: l'accesso a ChatGPT nell'ultimo anno è aumentato del 65% mentre il tempo speso è più che raddoppiato (118%).

La diffusione capillare dell'IA generativa si riverbera esplicitamente nel comportamento delle nuove generazioni. Secondo uno studio del 2025 della Deloitte, in Italia metà dei GenZ e quattro Millennial su dieci utilizzano la GenAI quotidianamente, con applicazioni che spaziano dalla creazione di contenuti all'analisi dei dati, dal project management allo sviluppo di software. Il 73% della Gen Z e il 73% dei Millennial italiani affermano che la GenAI ha liberato tempo e ha migliorato il work-life balance, mentre il 71% della Gen Z e il 76% dei Millennial pensa che abbia migliorato la qualità del proprio lavoro. Anche nel campo aziendale l'Italia ha un buon posizionamento rispetto ad altri Paesi europei. Il 53% delle grandi imprese ha acquistato licenze di strumenti di GenAI (principalmente ChatGPT o Microsoft Copilot), superando Francia, Germania e Regno Unito. Tuttavia, permane un divario significativo, probabilmente dovuto anche al lag temporale con il quale gli incrementi di produttività solitamente si disvelano nelle metriche di misurazione: il 39% delle grandi imprese che utilizzano questi strumenti ha riscontrato un effettivo aumento della produttività, mentre un ulteriore 48% non ha ancora valutato in modo quantitativo gli impatti.

L'entusiasmo per l'IA generativa si accompagna tuttavia al sorgere di altrettanto legittime inquietudini. Il 62% della Gen Z e il 67% dei Millennial considera opportunità di lavoro che siano meno vulnerabili all'automazione, mentre il 55% dei Millennial e il 61% dei Gen Z reputa che l'applicazione della AI comporta una riduzione dei posti di lavoro. La consapevolezza degli impatti profondi che l'IA sta già avendo e avrà via via di più nel mondo del lavoro rende ancor più urgente la necessità di sviluppare le competenze giuste per navigare questa trasformazione.

### **Quali sono le competenze chiave nell'era dell'intelligenza artificiale?**

Nel contesto attuale in rapida e profonda evoluzione tecnologica emerge una domanda cruciale: quali sono le competenze chiave richieste in una fase di trasformazione profonda dei modelli economici, culturali e organizzativi? L'ascesa della IA ci ingiunge a ripensare non soltanto le tecnologie, ma soprattutto le "capacità umane" fondamentali per governare le tecnologie intese quale momento di potenziamento delle capacità dell'essere umano e non invece come suoi possibili sostituti.

Nell'economia contraddistinta dalla AI, la creatività umana rimane centrale, amplificata dall'IA, ma sempre guidata da atteggiamenti critici, etici e progettuali.

### **IA: prolificità senza intenzionalità**

I sistemi di IA generativa si distinguono per la loro capacità di produrre contenuti in modo rapido e variegato, basandosi su modelli probabilistici e pattern ricorrenti. Tuttavia, essi non posseggono né intenzionalità e neppure metacognizione, ovvero le capacità umane di riflettere consapevolmente sul proprio operato e sul contesto in cui si inserisce l'output prodotto. Pertanto, il valore aggiunto resta nel giudizio e nelle capacità critica tipicamente umane, elementi che l'IA non può replicare autonomamente.

La prolificità della IA è un elemento sicuramente importante: essa può generare testi, immagini, codici e analisi in pochi secondi. Eppure, la velocità produttiva porta con sé il rischio di fomentare negli individui la dipendenza acritica, in cui l'abbondanza di output generati può offuscare la necessità di valutazione qualitativa. Infatti, di norma l'IA opera attraverso correlazioni statistiche apprese da enormi base dati, ma non possiede comprensione del significato, manca di corporeità, non ha esperienza vissuta, non conosce il contesto culturale specifico in cui si inserisce la sua produzione.

Sono queste limitazioni a definire il campo d'azione insostituibile dell'essere umano: la capacità di dar senso, direzione e significato all'output tecnologico. Se l'IA può suggerire mille variazioni di un testo o proporre soluzioni basate su pattern precedenti, solo l'intelligenza umana è in grado di valutare quale di queste opzioni sia più appropriata per un contesto specifico, quale risuoni con valori culturali condivisi, quale sia davvero innovativa rispetto a quanto è già noto.

### **Le cinque competenze chiave**

In tale scenario emergono cinque competenze fondamentali che rappresentano il cuore pulsante della nuova alfabetizzazione indispensabile per veleggiare nell'era dell'intelligenza artificiale:

#### **AI literacy**

Anzitutto comprendere come funziona l'intelligenza artificiale è un momento fondamentale: conoscere i modelli, i loro limiti, comprendere come impostare le istruzioni (i cosiddetti *prompt*) per ottenere risposte efficaci, e saper verificare fonti e dati. Questa alfabetizzazione di base sulla AI permette di utilizzarla come un partner didattico e professionale, valorizzandone le capacità senza perdere sensibilità e giudizio umano. La AI *literacy* include altresì la capacità di maneggiare e interpretare i dati, riconoscendone importanza e uso corretto nell'ambito personale, lavorativo e formativo. L'alfabetizzazione alla IA non richiede necessariamente competenze tecniche approfondite di programmazione o data science, ma implica una comprensione concettuale, che possiamo definire "semantica" delle modalità con cui questi sistemi apprendono, quali sono i loro punti di forza e quali i loro limiti tipici. Significa sapere che un modello linguistico può generare testo in modo corrente ma non necessariamente accurato, che può riprodurre bias presenti nei suoi dati di addestramento, che la sua conoscenza è limitata a un certo momento temporale. Include anche competenze molto pratiche come la conoscenza del *prompt engineering* e delle sue conseguenze: saper formulare richieste chiare, strutturate e contestualizzate per ottenere risultati più pertinenti e utili, evitando l'utilizzo becero e acritico del modello linguistico.

#### **Valutazione critica**

La capacità di leggere, sperimentare e rifinire i risultati prodotti dalla IA è una competenza essenziale. Non basta accettare passivamente le proposte generate dai sistemi; occorre mantenere sempre un approccio critico, evitando di incorrere in dipendenze e pregiudizi cognitivi che penalizzano l'emergere di idee diverse e innovative. Il pensiero critico consente di valutare, con la dovuta consapevolezza, qualità e affidabilità delle informazioni fornite, tutelando per tale via creatività e giudizio umano. La valutazione critica si traduce in una serie di pratiche concrete: la verifica delle fonti delle informazioni fornite dalla IA risalendo sempre alla prima fonte, il confronto delle varie risposte alla stessa domanda, il riconoscimento di quando l'output presenta bias o limitazioni, e soprattutto il mantenimento della capacità di dissentire dalle proposte della IA quando queste non corrispondono agli obiettivi o ai valori del progetto. Questa competenza richiede una forma di alfabetizzazione informativa più avanzata, che include la capacità di riconoscere allucinazioni, inconsistenze logiche e limiti contestuali nelle risposte generate.

### **Progettazione collaborativa**

L'approccio all'IA non è dunque di uso passivo e inconsapevole, ma di *co-creazione responsabile*. L'IA interviene nella genesi delle idee e nella riorganizzazione dei processi, eppure essa riesce a esprimersi meglio quando si inserisce in team dotati di capacità di orchestrare velocità e varietà delle sue proposte con il discernimento umano. Si affermano forme di "creatività distribuita" in cui ruoli tradizionali sono più flessibili e si lavora con gerarchie più piatte, coniugando governance, responsabilità e funzioni. La progettazione collaborativa con l'IA richiede un cambio di paradigma mentale rispetto ai modelli di lavoro tradizionali. Non si tratta più di delegare compiti specifici a uno strumento che li esegue meccanicamente, ma di instaurare un *dialogo iterativo* in cui l'essere umano guida, l'IA propone, l'essere umano valuta e raffina, in un ciclo di miglioramento continuo. Questo approccio esige competenze avanzate di coordinamento, la capacità di integrare contributi umani e artificiali in modo armonico, e una chiara definizione delle responsabilità finali, che non possono che restare in capo all'essere umano.

### **Differenziazione e originalità**

Un rischio tecnologico che l'AI pone si riferisce alla cosiddetta "omologazione", per l'intervento di modelli e fonti condivisi che portano a risultati simili. In un mondo dove centinaia di milioni di persone utilizzano gli stessi strumenti di IA generativa, alimentati dagli stessi dataset e guidati da algoritmi simili, il rischio di convergenza creativa è reale e tangibile. Se tutti usano ChatGPT per scrivere, Midjourney per creare immagini, e gli stessi prompt circolano online, il risultato non può che essere una progressiva standardizzazione dell'output creativo. La capacità di differenziazione, ovvero la capacità di sviluppare un linguaggio personale, di portare prospettive uniche, di attingere a riferimenti culturali specifici, di coltivare un gusto distintivo che non può essere replicato dall'IA, diviene quindi una competenza strategica fondamentale. La vera sfida competitiva è dunque saper conservare la *diversità creativa* curando la propria impronta distintiva e le proprie capacità estetiche. Nei settori creativi questo indica lo spostamento verso profili con competenze ibride, capaci di fondere competenze tradizionali e nuove expertise algoritmiche.

### **Etica operativa**

Affrontare questioni come la proprietà intellettuale, la paternità delle creazioni, la trasparenza degli strumenti e la qualità dei dati non è solo un esercizio teorico, ma una pratica quotidiana. La governance della IA è una competenza trasversale che si manifesta in regole di policy e in strumenti manageriali verificabili, che garantiscono trasparenza e responsabilità sociale nelle applicazioni

dell'intelligenza artificiale. L'etica operativa trascende i principi astratti per tradursi in decisioni concrete che ogni professionista deve affrontare quotidianamente: è etico utilizzare contenuti generati dall'IA senza dichiararlo? Come si fa a gestire la proprietà intellettuale quando un'opera è co-creata con l'IA? Quali responsabilità abbiamo quando i sistemi che utilizziamo possono perpetuare discriminazioni o bias? Come garantire che l'uso della IA nella propria organizzazione sia trasparente e affidabile? Le risposte a queste domande richiedono non solo sensibilità etica, ma anche la capacità di tradurre valori in protocolli operativi, linee guida chiare e meccanismi di verifica.

### **La competenza trasversale della strategia consapevole**

Oltre alle cinque competenze sin qui definite, diviene essenziale una sesta competenza, questa volta olistica e comprensiva, ovvero la capacità di sviluppare una *visione integrata*, una strategia in grado di sussumere tutte le dimensioni dell'interazione con l'IA. La strategia consapevole di approccio alla AI rappresenta una meta-competenza che permette di navigare la complessità della IA senza perdere di vista l'obiettivo finale, ossia l'utilizzo della tecnologia per potenziare le capacità umane e non per sostituirle o ridurle. Tale approccio richiede la capacità di trascendere il facile entusiasmo per le novità tecnologiche dei tecno-ottimisti per poter antivedere i loro sviluppi e mantenere una prospettiva di lungo termine sugli impatti culturali, sociali ed economici della IA. Tale visione strategica favorisce la salvaguardia del controllo umano sulla tecnologia, valorizzandone il potenziale senza dimenticare la natura intenzionale e metacognitiva propria degli esseri umani, che la IA non è in grado di replicare. In tal modo, l'AI diviene un alleato capace di rinvigorire e potenziare sotto la guida sapiente di competenze umane consapevoli e responsabili.

### **Sfide e opportunità per il lavoro e la formazione**

Com'è noto, l'evoluzione tecnologica richiama lo sviluppo di profili professionali ibridi e multidisciplinari, capaci di integrare competenze umanistiche, tecniche, critiche ed etiche. I dati mostrano chiaramente questa tendenza: le imprese cercano via via di più professionisti che non siano soltanto esperti tecnici o creativi, ma che sappiano muoversi in maniera fluida fra questi domini, traducendo illuminazioni tecnologiche in soluzioni umane e viceversa.

Nel mondo del lavoro e della formazione, l'introduzione della AI non è soltanto una risorsa, ma altresì una vera sfida educativa: formare una cittadinanza digitale attenta, empatica e capace di condurre un dialogo proficuo con le tecnologie emergenti. Le istituzioni educative sono pertanto chiamate a ripensare i loro progetti formativi per includere non soltanto competenze tecniche sulla IA, ma anche capacità critiche di valutazione, sensibilità etica e abilità collaborative. La formazione continua (*lifelong learning and teaching*) diviene essenziale e ineludibile, poiché le tecnologie evolvono rapidamente e le competenze che vanno bene oggi non saranno certamente adeguate entro pochi mesi.

Per questo motivo, le competenze cognitive e sociali, quali intelligenza emotiva, comunicazione efficace, collaborazione e gestione dei conflitti assumono un ruolo centrale in contesti dove essere umano e macchina debbono cooperare per raggiungere risultati più elevati. Tali *soft skills*, un tempo considerate complementari rispetto alle competenze tecniche, assumono un ruolo centrale e distintivo dacché rappresentano l'area in cui l'intelligenza umana mantiene ancora un vantaggio rispetto alla IA. Emerge inoltre la necessità di sviluppare resilienza e adattabilità: la capacità di apprendere continuamente, di disimparare quando necessario, di rimanere flessibili di fronte al cambiamento. In un contesto dove le competenze tecniche divengono obsolete nel giro di pochi mesi

o anni, la capacità di *apprendere ad apprendere* diventa la competenza fondamentale che sostiene tutte le altre.

## Conclusione

Le competenze decisive nell'era della AI sono radicate nelle dimensioni umana, critica e strategica. I dati mostrano una diffusione rapidissima e capillare della IA generativa, con centinaia di milioni di utilizzatori che la utilizzano nelle loro pratiche quotidiane, professionali e creative. Eppure, quasi paradossalmente la diffusione capillare e pervasiva della AI non riduce, ma anzi accentua, la necessità di competenze umane distintive. L'AI non è un sostituto, ma un amplificatore della creatività e della produttività umana che richiede un approccio consapevole e multidimensionale. Le cinque competenze chiave – AI literacy, capacità di valutazione critica, progettazione collaborativa, differenziazione e originalità, etica operativa – insieme alla capacità di sviluppare una strategia consapevole olistica, rappresentano i cinque pilastri fondamentali sui quali è possibile edificare una relazione produttiva e sostenibile con l'IA.

Investire in tali competenze significa saper affrontare con efficacia le sfide contemporanee, generare nuove opportunità di crescita e di innovazione, e disegnare un approccio economico umano, etico e sostenibile in un mondo sempre via via più digitalizzato. Non si tratta di resistere al cambiamento tecnologico, né di abbracciarlo acriticamente, ma di guidarlo con saggezza, mantenendo al centro i valori umani che danno significato al progresso. In conclusione, l'era dell'intelligenza artificiale non è l'era della sostituzione dell'umano, ma l'era della sua riaffermazione: proprio quando le macchine diventano sempre più capaci di fare, le qualità umane distintive – creatività, giudizio etico, sensibilità contestuale, capacità di dar senso – divengono ancora più preziose e difficilmente sostituibili.

## Bibliografia

- AI for K12 Initiative (2021). *AI4K12 Five Big Ideas in AI*. Association for the Advancement of Artificial Intelligence & Computer Science Teachers Association. <https://ai4k12.org>
- Boden, M. A. (2010). *The Creative Mind: Myths and Mechanisms* (2nd ed.). London: Routledge.
- Chiriatti, M. (2022). *Incoscienza artificiale: Come fanno le macchine a prevedere per noi*. Egea, Milano.
- Deloitte (2025). *GenZ and Millennial Survey 2025*. Deloitte.
- D'Amora, R. (2024). AI e competenze umane: verso l'intelligenza centaurica. *Agenda Digitale*.
- European Commission (2019). *Ethics Guidelines for Trustworthy AI*. High-Level Expert Group on Artificial Intelligence.
- Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*, 1(1).
- Giray, L. (2023). Prompt engineering with ChatGPT: A guide for academic writers. *Annals of Biomedical Engineering*, 51(12), 2629-2633.
- ISTAT (2024). *Imprese e ICT 2024*. Rome Istituto Nazionale di Statistica.
- Jobin, A., lenca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399.

Korzynski, P., Mazurek, G., Krzypkowska, P., & Kurasinski, A. (2023). Artificial intelligence prompt engineering as a new digital competence: Analysis of generative AI technologies such as ChatGPT. *Entrepreneurial Business and Economics Review*, 11(3), 25-37.

Laupichler, M. C., Aster, A., Haenlein, M., & Raupach, T. (2023). Development of the “Scale for the assessment of non-experts' AI literacy” – An exploratory factor analysis. *Computers in Human Behavior Reports*, 12, 100338.

Lee, K.-F., & Qiufan, C. (2021). *AI 2041: Ten Visions for Our Future*. Currency, New York.

Long, D., & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI '20), 331, 1-16.

Lu, J., Zheng, R., Gong, Z., & Xu, H. (2024). Supporting Teachers' Professional Development With Generative AI: The Effects on Higher Order Thinking and Self-Efficacy. *IEEE Transactions on Learning Technologies*, 17, 1267-1280.

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041.

Osservatori Digital Innovation, (2024). *Artificial Intelligence Market Report 2024*. Politecnico di Milano, Milano.

Panciroli, C., Corazza, L., & Reggiani, A. (2020). AI Literacy e competenze per l'intelligenza artificiale. *Media Education*, 11(2), 145-162.

Prencipe, A., & Sideri, M. (2023). *Il visconte cibernetico: Calvino e l'intelligenza artificiale*. LUISS University Press, Roma.

Roncaglia, G. (2024). *L'architetto e l'oracolo: Forme della conoscenza nell'era digitale*. Laterza, Bari.

Shneiderman, B. (2022). *Human-Centered AI*. Oxford University Press.

World Economic Forum (2023). *The Future of Jobs Report 2023*. World Economic Forum, Geneva.

## **Giornalismo e verità: chi controlla il racconto controlla il reale - di Francesca Rizzuto, Università degli studi di Palermo**

*Le piattaforme digitali ridefiniscono l'informazione, favorendo mini-narrazioni individuali che sfidano i modelli giornalistici tradizionali e legittimano la coesistenza di verità multiple in un contesto sempre più polarizzato e frammentato*

*Non c'è nulla al mondo più forte di una buona storia. Niente può fermarla, nessun nemico può sconfiggerla perché tra i poteri del racconto c'è anche quello di ingannare (Brooks, 2022, 11).*

Secondo Brooks (2022), dopo la fine delle “grandi narrazioni” nell’ecosistema mediale ibrido contemporaneo siamo sedotti dalle storie e nei social ci ritroviamo “inondati da una moltitudine di mini-narrazioni, individuali o collettive e, in molti casi, narcisistiche” (Brooks, 2022, p.13); tale proliferazione pervasiva e planetaria di racconti è evidente anche in ambiti diversi dai media digitali, come in quello politico o in altre istituzioni sociali, al punto che il discorso pubblico, che dovrebbe essere basato sull’analisi razionale, è ormai dominato dalla iperinflazione delle storie di racconti, diventate armi cruciali per governare e per comprendere l’ambiente circostante. Nella società delle piattaforme (Van Dijck, Poell, & De Waal, 2018), la narrazione sembra essere definitivamente accettata come l’unica forma di conoscenza, via indispensabile per la comprensione della realtà (Brooks, op. cit, pp.11-13) ed è usata dai leader politici come strategia retorica efficace per trasmettere i propri contenuti e per controllare le narrazioni dei media. L’esito della conquista definitiva messa in atto dallo storytelling si rivela nella presenza di un *nuovo ordine narrativo* (Salmon, 2008) connesso ad un incessante processo di “costruzione narrativa della realtà” che, nell’era dei social media pervasivi e planetari, assume i contorni di una generale tendenza al *controllo narrativo dell’esperienza*, radicalmente diverso dalla presenza delle narrazioni, che hanno caratterizzato la storia umana anche in epoche pre-mediali. Se è indubbio che, come sostenuuto da Barthes, “non è mai esistito in alcun luogo un popolo senza racconti” (Barthes, 1969, trad. it. p.7), in anni recenti l’espressione di Niles (2010), *homo narrans* ha evidenziato la centralità del racconto come elemento costitutivo dell’esistenza umana: le storie hanno sempre rappresentato una parte indispensabile dell’interazione tra un individuo ed il mondo circostante fin dall’infanzia (Bruner, 1991); del resto, è possibile sapere cosa è accaduto attraverso le versioni raccontate dei fatti, vale dire quando gli eventi sono trasformati in *storie* (Brooks, op. cit., p.18), come ha scientificamente confermato anche la psicoanalisi di Freud, per la quale la via per la comprensione del sé è l’uso di sequenze narrative grazie alle quali è possibile individuare cause e origini dei percorsi psicologici individuali. Tuttavia, rispetto ai contesti predigitali, la centralità odierna della narrazione svela alcuni tratti peculiari del clima culturale contemporaneo della post-verità (Corner ,2017; McIntyre, 2018; Joux, 2023), nel quale dilaga una frattalizzazione della verità in una produzione incessante di racconti individuali ed emotivamente coinvolgenti sulla realtà, legittimando verità multiple che circolano a livello planetario, e all’interno di bolle auto-isolanti, consentono pratiche di interazione caratterizzate da tendenze all’omofilia, all’evitamento della dissonanza ed al *confirmation bias*. In tale contesto, più attento all’efficacia emotiva che alla oggettività dei fatti, i social media incoraggiano pratiche di storytelling in cui soggetti differenti (individui, leader politici, istituzioni, aziende) sono quotidianamente impegnati nella produzione di contenuti informativi frammentati e svincolati dalla mediazione professionale e dai limiti testuali dei formati giornalistici tradizionali. Nelle pagine seguenti l’attenzione verrà focalizzata

sulle ricadute della nuova centralità della narrazione sul giornalismo, al quale spetta il compito di selezionare e raccontare fatti in una fase storica dominata da un marcato processo di *storificazione della realtà*, per cui tutti gli eventi entrano nell'orbita della narratività e gli individui agiscono e si percepiscono esclusivamente in un mondo di "storie". Nella società delle piattaforme sono favorite pratiche di narrazione individuale, si offrono spazi inediti di democratizzazione del privato, di autorappresentazione (Castells, 2009) e di auto-narrazione, nei quali "la notizia è sempre meno pensata come documento e sempre più come "racconto" che punta molto sulla proiezione individuale, sull'identificazione e sull'appassionamento" (Lorusso, 2018, 119). Erede di rilevanti processi già avviatisi nei decenni scorsi (come la spettacolarizzazione della realtà o l'accentuazione della componente emotiva per attrarre destinatari nel contesto della crescente competizione nel mercato editoriale in forte crisi economica), l'informazione prodotta *nei social e per i social* propone, pertanto, testi giornalistici intessuti da logiche narrative, che si inseriscono in sceneggiature condivise: prende così forma un sistema granulare di news, attento più all'efficacia narrativa che all'attendibilità, che presenta, nella forma della plausibilità, vale a dire del verosimile, verità "opache" ma perfettamente in grado di diventare oggetto di storytelling (Groot Komerlink, Costera Meijer 2015).

Per comprendere la nuova rilevanza del processo di *costruzione narrativa della realtà* e le sue conseguenze per la professione informativa, il punto di partenza più utile è la premessa della differenza basilare tra ciò che è accaduto (l'evento) e il *racconto* dell'evento, vale a dire la "tra realtà e la sua rappresentazione", distinzione proposta dai formalisti russi come dicotomia tra intreccio e fabula: "la fabula corrisponde agli avvenimenti narrati nel loro ordine cronologico naturale, mentre l'intreccio è la presentazione degli eventi nella narrazione che leggiamo o ascoltiamo, eventi che possono non rientrare nell'ordine cronologico, che possono essere riordinati, ingranditi, minimizzati, distorti. Quindi, l'intreccio non è innocente: è una ripresa di una prospettiva, una disposizione" (Brooks, op. cit, 21). Tale distinzione ha rappresentato il pilastro fondamentale per l'evoluzione del giornalismo il cui destinatario è, sin dalle origini secentesche, il lettore nato nell'era gutemberghiana, membro di un pubblico senza luogo per il quale è giunto al culmine il processo di sgretolamento della linearità rigida del libro/testo scritto e di declino del carattere sociale dell'esperienza condivisa della lettura ad alta voce, che già si era avviato prima della stampa. Con il libro moderno il testo è progressivamente diventato accessibile, nel tempo e nello spazio, ad un *pubblico di lettori* non fisicamente compresenti, costituendo la base della lettura silenziosa e del ragionamento individuale, tratti tipici dell'esperienza testuale sequenziale propria del giornalismo moderno (Cappello, Rizzato, 2024). Un momento di svolta epocale nel rapporto tra narrazione e realtà è riconducibile all'affermazione del romanzo borghese settecentesco, con il quale viene segnata la "distinzione tra realtà (esterna) e finzione (interna al testo), cioè tra *realità reale* nella sua quotidianità e *messa in scena* del reale, grazie alla produzione di un mondo parallelo separato e, al contempo, connesso nelle forme della narrazione» (Boccia Artieri 2006: 192). Questo genere letterario punta a raccontare il mondo e costruisce una *realità artificiale*, che si contrappone alla *realità vera*, creando un circuito comunicativo in cui i lettori si possono identificare con storie e personaggi, e, quindi, accedono al mondo reale dei sentimenti umani attraverso la narrazione di vite "non reali". La narrazione romanzesca offre un allargamento dello spazio sociale rappresentato, che ha avuto enormi ricadute sul processo di legittimazione sociale del giornalismo, riconosciuto come ambito autonomo caratterizzato da un sempre più solido patto fiduciario tra giornalista/narratore/fonte e destinatari, rafforzato dalla periodicità che crea relazioni permanenti tra fonti e lettori. Nel caso dell'informazione, infatti, a differenza del romanzo, la questione della buona fede e dell'affidabilità del narratore diventa cruciale e costitutiva e i lettori devono distinguere ciò che è degno di fiducia da ciò che è sospetto, poiché nei giornali non può esserci l'illusione romanzesca, vale a dire quella "sospensione volontaria

dell'incredulità con cui i lettori si avvicinano ai romanzi, non necessariamente per ingenuità o per ottusità, ma perché questo fa parte del piacere intellettuale ed emotivo del leggere un romanzo” (Brooks, op. cit, pp. 36-37), il cui destinatario sa che si tratta di una finzione, ma accetta di sottostare alla simulazione del reale e di credere che la storia sia vera. L'attenzione alle questioni epistemologiche della narrazione e alla sua relazione con fatti reali è, quindi, stata centrale nel giornalismo: al contrario delle narrazioni fictionali, per la scrittura giornalistica si sono così delineate due direttive di sviluppo, che hanno dato vita a modelli narrativi radicalmente differenti, anche se sempre in un intreccio complesso di interazioni conflittuali e scambi (Broesma, 2007): da un lato, *l'information model*, per il quale l'aderenza rigida ai fatti è stata la base costitutiva della professione informativa, dall'altro, lo *story model*, che ha mirato soprattutto all'ampliamento del pubblico, proponendo storie attraenti e accattivanti costruite grazie al ricorso a tecniche narrative più vicine al racconto romanzesco (Cappello, Rizzato 2024).

### **La narrazione giornalistica nella platform society. Dalla crisi della piramide invertita alle esperienze testuali digitali.**

Accanto al rapporto fondativo con la realtà, l'evoluzione della narrazione giornalistica è stata segnata anche dai radicali cambiamenti delle tecnologie della stampa (Fliechy 1995), che in un percorso plurisecolare hanno imposto differenti modelli di esperienza testuale, compresi nell'universo informativo fino al Novecento, pur con significative discrasie o accelerazioni, che con i media digitali sono giunti alla totale ibridazione: il primo modello è stato definito da Eugeni (1996) “centripeto e sequenziale” ed è quello del racconto scritto o del film, caratterizzato dalla progressiva smaterializzazione del testo e dalla scarsa visibilità della fruizione; il secondo “policentrico e reticolare” è quello della stampa periodica o dei manifesti pubblicitari, caratterizzati da spazialità grafica e da una precisa griglia progettuale. Un posto centrale, inoltre, anche per la storia del giornalismo, è stato occupato dall'uso delle immagini stampate che dalle illustrazioni ottocentesche è arrivato fino al fumetto e alla fotografia, dando vita a forme miste di testualità, in cui la prospettiva inserisce lo spettatore nel testo, diventa matrice di tragitti interpretativi, struttura lo spazio e impone un potere ordinatore esterno alla storia, che delinea la grammatica (logico-sequenziale) della narrazione. Nell'ecosistema digitale ibrido le pratiche narrative con le quali gli individui presentano e condividono continuamente emozioni ed esperienze hanno prodotto conseguenze rilevanti sia per le modalità di interazione che per il giornalismo, imponendo una radicale ridefinizione del suo *campo* e dei suoi *oggetti*, così come una riflessione sul significato del ruolo sociale tradizionalmente attribuito al giornalista, cioè quello di agente di mediazione per i fruitori, a cui proporre quotidianamente una lettura degli eventi più rilevanti, utile per la loro partecipazione consapevole al dibattito pubblico. Rispetto alla sfera pubblica dell'era predigitale i giornalisti oggi selezionano e raccontano i fatti inserendosi in un dibattito pubblico in cui le piattaforme favoriscono circuiti di polarizzazione ideologica e processi di saturazione comunicativa, nei quali aumenta il rischio di disinformazione, fino al rischio di privilegiare l'adozione della menzogna come elemento strutturale della comunicazione (Sorice, 2020, pp.382-383). In tale contesto, se è indubbio che le tecnologie digitali hanno trasformato il modo di produrre e diffondere le notizie, segnando una crisi inarrestabile del giornalismo cartaceo tradizionale, diventa rilevante comprendere come e quanto alla rivoluzione delle routine produttive e della fruizione delle informazioni sia riconducibile la *crisi della fiducia* nel prodotto giornalistico e alla sua capacità di continuare a raccontare la realtà sulla base della dicotomia vero/falso, che ne è stata per secoli il parametro costituivo (Groot Kormelink, Costera Meijer, 2015; Sorrentino, Splendore, 2022). Rispetto al supporto cartaceo tradizionale, il mezzo televisivo aveva già avviato un processo di

commistione tra i criteri referenziali del *newsmaking* e gli imperativi commerciali della logica infoinfrattenitiva: nella fase pre-digitale si era realizzata vera e propria contaminazione reciproca, sia sul piano delle tecnologie che su quello delle strategie narrative, con la conseguenza che si sono prodotte molteplici forme di ibridazione tra modelli professionali differenti fino alla definitiva affermazione di un *giornalismo anfibio*, che racconta frammenti di realtà servendosi di un linguaggio visuale teatralizzante. In altri termini, la tv ha rivoluzionato gli stili narrativi per attrarre audience distratte con la conseguenza che il giornalismo televisivo ha abbandonato quasi del tutto lo stile dell'argomentazione razionale e della “verità” come obiettivo desiderabile per il professionista dell'informazione. Già alla fine degli anni Novanta, pertanto, l'idea del giornalismo come sistema altamente standardizzato di produzione di resoconti obiettivi dei fatti, ma anche come luogo del confronto tra idee, sembra essere regredita alla logica darwinista della pura e semplice contrapposizione emotiva, dell'obbligatorietà dello schieramento manicheo, della ricerca di facili capri espiatori, con esigue capacità di approfondimento in news, che ricostruendo i fatti all'interno di nuovi formati e generi tv, “narrano” il mondo in pochi secondi, presentando semplici capsule informative, nelle quali sono sempre più usati gli stilemi del *racconto fictional* (Rizzuto, 2023). Nell'era digitale, i continui cambiamenti nel comportamento dei consumatori e la crescente migrazione del pubblico verso le notizie online, oggi fonti primarie per l'accesso all'informazione, hanno imposto una radicale innovazione nel modo in cui si raccontano le notizie e sui criteri di notiziabilità che ne determinano i contenuti (Pavlik, 2000; Giles, 2004). Secondo Pavlik, le narrazioni giornalistiche proposte in ambiente digitale tendono a definire un nuovo tipo di *storytelling* giornalistico definito *contextualized journalism*, che presenta cinque dimensioni o aspetti basilari: ampiezza delle modalità comunicative, ipermedialità, coinvolgimento del pubblico, contenuti dinamici *customization*. Come evidenziato da Zaganelli e Marino (2016) il giornalismo on line può raccontare storie utilizzando differenti codici con i quali costruire un *racconto arricchito*, in cui il lettore può essere coinvolto da più punti di vista sensoriali oltre che nella decodificazione cognitiva della narrazione, ad esempio, potenziando le narrazioni visive. Una delle caratteristiche più rivoluzionarie del giornalismo digitale è l'ipermedialità, che associa all'arricchimento mediale integrato su un unico canale comunicativo, quello digitale, un arricchimento semantico creato per mezzo di link che connettono le notizie e rendono possibile un percorso di lettura tematico, oggi confluito nella produzione di contenuti crossmediali. Naturalmente, la peculiarità principale dell'informazione on line è il nuovo ruolo dei destinatari: rispetto al giornalismo della carta stampata, lo *storytelling* immersivo degli ambienti digitali, garantisce al singolo utente la possibilità di interagire con la notizia a più livelli, dalla semplice selezione delle preferenze di lettura all'integrazione di informazioni (Zaganelli e Marino, 2016, p. 99). In un contesto in cui la notizia è diventata una monade fluida e dinamica, essa può essere continuamente esposta ad aggiornamenti che la rendono sempre una notizia in tempo reale: la conseguenza più significativa è che “la narrazione delle notizie si sovrappone al flusso delle vita e non è più solo uno snodo informativo che precede le azioni, ma le accompagna costantemente” (ibidem).

Infine, la presenza nelle redazioni di un marcato orientamento al cliente, che si concretizza nella personalizzazione del flusso informativo nel suo rapporto con il lettore, che può scegliere le notizie e il modo di associarle, oltre che richiedere contenuti particolareggiati e approfondimenti dettati da interessi personali. La crescente customizzazione delle news ha determinato una svolta significativa nelle pratiche della scrittura giornalistica, vale a dire la messa in discussione del sistema a "piramide rovesciata", che organizza i fatti in ordine decrescente di importanza e che è stato per decenni uno dei pilastri della pratica giornalistica e dello *storytelling* (Norambuena et al., 2020). Anche se, come osservano Kulkarni et al. (2024) questo metodo di scrittura è ancora predominante per scrivere storie di "hard news", nella platform society i nuovi media consentono una miriade di nuove modalità di

narrazione, in cui l'istanza narrativa e quella ricettiva, anche se di fatto non coincidono, sono immerse nella rapidità del flusso comunicativo e sembrano sovrapporsi, creando possibilità inedite di coinvolgimento tra utenti e notizie. La questione delle strategie narrative del giornalismo digitale è al centro di numerosi studi che focalizzano l'attenzione sul declino recente dell'attenzione dei pubblici rispetto alle forme tradizionali di storytelling (soprattutto tra le fasce più giovani), e si chiedono se, nell'era digitale e dei social media, esistano modi alternativi (*più informativi e coinvolgenti*) che i professionisti dell'informazione possono usare per raccontare le notizie. A dire il vero, la storia del giornalismo è stata da sempre attraversata dalla ricerca costante di nuove forme potenzialmente più coinvolgenti di storytelling informativo, in quanto sebbene le notizie possano attrarre lettori, spettatori e ascoltatori per il loro contenuto, il modo in cui tali contenuti vengono presentati spesso ne determina l'impatto. Secondo Canavilhas (2007), l'affermazione della narrazione a "piramide rovesciata" è emblematica e può essere ricondotta all'inizio del XIX secolo, periodo durante il quale i giornalisti inviavano le notizie servendosi del telegrafo, tecnologia che gli imponeva di strutturare i testi in paragrafi, obbedendo all'imperativo di comunicare i fatti in ordine di importanza (Pöttker, 2003). Da questo momento si è così affermato un tipo molto particolare di gerarchia informativa, in netto contrasto con la struttura narrativa cronologica, in cui gli eventi sono sistematicamente riportati nell'ordine in cui si sono verificati e dove il culmine della storia giunge alla fine (Norambuena et al., 2020; Thomson et al., 2008; Walters, 2017). Dal punto di vista produttivo e organizzativo, questo modello si è rivelato vantaggioso per la stampa, perché ha consentito ai redattori di tagliare i testi dal basso verso l'alto senza perdere gli elementi chiave della storia (Pöttker, 2003). Inoltre, con l'informazione prioritaria rispetto alla suspense (Van Krieken, 2018; Van Krieken, Sanders, 2019), la piramide rovesciata ha reso le notizie più veloci da scrivere e più veloci da leggere: successivamente, in seguito alla ridefinizione del giornale come bene di consumo quotidiano, tale schema di scrittura giornalistica si è affermato come il modello più efficace ed economico per la comunicazione delle notizie (DeAngelo e Yegian, 2019), non solo per il giornalismo cartaceo, ma anche per quello radiotelevisivo, e, molto più tardi, è stato adottato dalle notizie online sui nuovi media. In generale, il successo della narrazione a *piramide rovesciata* può essere compreso prendendo in considerazione il fatto che essa si adatta bene alla crescente enfasi sulle ultime notizie, vale a dire sull'obbligo dell'aggiornamento: con il tempo, però, è stata erroneamente associata all'obiettività, con la conseguenza che viene data rilevanza sempre più ai fatti e meno all'interpretazione giornalistica, facendo prevalere l'idea che le narrazioni lineari significhino un resoconto "meno oggettivo" nonostante l'evoluzione di ogni evento sia sempre logicamente cronologica. Nell'era delle capsule informative prodotte e diffuse nei social, il predominio della narrazione a "piramide rovesciata" nel giornalismo rimane significativo, radicato in routine e forme consolidate, ma la sua efficacia come forma di comunicazione è sempre più messa in discussione, tanto che, per alcuni, è diventata "una specie di dinosauro" (Yaros, 2006 p. 287). Secondo Ward, infatti, questo modello implica comunque "un intervento interpretativo più strutturato rispetto alla narrazione lineare, in cui i giornalisti impongono necessariamente una gerarchia di importanza fattuale nel racconto di storie in divenire". (Ward, 2009 p. 299). La rigidità delle strutture narrative tradizionali si è scontrata con quella online e molti studi recenti si sono concentrati sui modi in cui gli elementi strutturali delle notizie possono influenzare il ricordo, la comprensione e la percezione, vale a dire sull'impatto della narrazione riguardo ai livelli di comprensione e coinvolgimento del pubblico, esplorando anche altre forme di narrazione, come ad esempio la crescente sovrapposizione tra i generi giornalistici e lo stile comico/satirico dei meme. Mettere in discussione la rilevanza dei formati giornalistici consolidati rimane fondamentale, soprattutto nel contesto del calo del pubblico per la maggior parte dei formati giornalistici tradizionali: la riflessione scientifica sulle strutture narrative all'interno delle notizie

sembra supportare l'affermazione gli sviluppi del giornalismo online abbiano messo in crisi il predominio del modello della piramide rovesciata e che si stia diffondendo un giornalismo narrativo che fa ricorso a "stili narrativi più creativi e meno strutturati" (Johnston, 2017 p. 14) ricercando modi per diffondere notizie e informazioni più efficaci e accessibili al pubblico. La rivoluzione digitale ha pertanto trasformato il campo giornalistico, favorendo il passaggio dai tradizionali processi sequenziali (selezione, verifica, gerarchizzazione, presentazione, fruizione) ad un intreccio magmatico tra produzione, distribuzione e consumo, a cui è riconducibile una nuova declinazione della compenetrazione tra logiche di intrattenimento e di partecipazione, che, dando maggiore rilevanza alla componente emotiva (Wahl-Jorgensen, 2020), contribuisce al declino di credibilità degli attori tradizionali della mediazione, come fonti autorevoli e affidabili di verità. Il modo in cui le notizie vengono prodotte e consumate nell'era digitale ha reso più opaco il confine tra *producer* e *consumer*, con la conseguenza che il linguaggio dell'informazione sta diventando sempre più multicanale, all'interno di una narrativa con testualità iper-frammentata, nelle quali la perfetta aderenza ai fatti, viene relegata in posizione secondaria rispetto alla dimensione emozionale e la verità diventa un ambito di scontro tra diverse visioni del mondo. Il modello orientato alla legittimità di tutte le opinioni si impone sempre più anche nella logica editoriale dell'informazione che trasforma la realtà in una controversia, rispetto alla quale occorre elaborare scelte di posizionamento: la conseguenza più significativa è un giornalismo che presenta tutte le versioni di una storia, con il rischio di "dare visibilità a contro-narrazioni che sono false, ma che comunque competono in un agone, che finisce per essere più narrativo che informativo" (Bentivegna, Boccia Artieri, 2021 p. 118).

### Osservazioni conclusive

Nell'ambiente digitale le regole che determinano la costruzione dell'involucro narrativo vengono ridiscusse alla luce di un'istanza che prevede la partecipazione del lettore in questo processo.

Negli ultimi anni, la digitalizzazione ha determinato una radicale *complessificazione delle dinamiche informative*: nel sistema convergente e ibrido contemporaneo è stata possibile la deterritorializzazione dell'informazione, che ha reso difficile sia l'individuazione degli attori sociali coinvolti nella produzione e diffusione di notizie, che le loro finalità, lasciando intravedere il potenziale collasso della funzione di *gatekeeping* da parte dei professionisti dell'informazione. In un contesto mediale in cui tecnologie del passato si trovano accanto ai nuovi strumenti, l'esigenza di essere informati o di far circolare idee può essere oggi soddisfatta "da un'offerta che non si esaurisce in quella costruita dagli attori tradizionali della mediazione informativa ma coincide addirittura con l'ambiente mediale nel suo complesso" (Bentivegna, Boccia Artieri, *op. cit.*, p. 40). Il lettore è sempre più interessato al racconto e sempre meno alle sue fonti. Più che tradurre la realtà, "la narrazione agisce nella realtà: stimola mode, crea tendenze, detta abitudini di consumo, favorisce atteggiamenti cognitivi e passionali. Ed è il controllo del racconto, e non delle sue fonti, che determina il controllo dell'impatto sociale di ogni narrazione" (Zaganelli, Marino, *op. cit.*, p. 93). Le pratiche interattive, grazie alle quali la notizia può essere smontata e rimontata per venire approfondita da molteplici punti di vista (Usher, 2014), così come le tendenze alla narrazione estesa e alla multiauthorialità, tipiche della transmedialità (McErlean 2018) impongono alla ricerca di analizzare il ruolo dei nuovi fornitori di notizie e trasmettitori di eventi (García-Orosa, López-García, Vázquez-Herrero 2020), soprattutto al tempo in cui le tecnologie di intelligenza artificiale, sono ormai perfettamente capaci di produrre costruzioni narrative fluide e verosimili, che si inseriscono velocemente in dinamiche virali planetarie.

## Bibliografia

- Barthes R. et al. (1969), *L'analisi del racconto*, Bompiani, Milano.
- Bentivegna S., Boccia Artieri G. (2021), *Voci della democrazia. Il futuro del dibattito pubblico*, Il Mulino, Bologna.
- Boccia Artieri G. (2006), “Farsi media. Consumo e media-mondo tra identità, esperienze e forme espressive”, in Di Nallo E. e Paltrinieri R., *Cum Sumo. Prospettive di analisi del consumo nella società globale*, Franco Angeli, Milano.
- Broersma M., (2007), *Form and Style in Journalism: European Newspapers and the Representation of News, 1880-2005*, Peeters, Lreuen.
- Brooks P. (2022), *Sedotti dalle storie. Usi e abusi della narrazione*, Carocci, Roma.
- Bruner, (1991), The Narrative Construction of Reality Bruner, *Jerome Critical Inquiry*, 18, 1-20.
- Canavilhas, J. 2007. “Web Journalism: From the Inverted Pyramid to the Tumbled Pyramid.” Biblioteca on-line de ciências da comunicação.
- Cappello G., Rizzato, *Giornalismo, post-verità e media education*, (2024), Mimesis, Milano.
- Castells, M. (2009). *Communication and Power*, Oxford University Press Ox- ford. (Trad. it.: *Comunicazione e potere*. Egea, Milano, 2009).
- Chadwick A. (2013), *The hybrid media system. Politics and power*, Oxford University Press, Oxford-New York.
- Chambers S. (2021), “Truth, Deliberative Democracy, and the Virtues of Ac- curacy: Is Fake News Destroying the Public Sphere?”, *Political Studies*, Vol. 69(1) 147-163, Sage, London.
- Corner J. (2017), Fake news, post-truth and media-political change, in *Media, Culture & Society*, 39 (7), 1100-1107.
- DeAngelo, T. I., and N. S. Yegian. 2019. “Looking for Efficiency: How Online News Structure and Emotional Tone Influence Processing Time and Memory.” *Journalism & Mass Communication Quarterly* 96 (2): 385–405.
- Eugenio R. (1996), *Il testo visibile. Teoria, storia e modelli di analisi*, Carocci, Roma.
- Flichy, p. (1995), *Storia della comunicazione moderna. Sfera pubblica e di- mensione privata*, Baskerville, Bologna.
- Giles, B. 2004. “Thinking About Storytelling and Narrative Journalism.” *Nieman Reports* 58 (1): 3.
- Groot Kormelink T., Costera Meijer I. (2015), Truthful or Engaging? Surpassing the dilemma of reality versus storytelling in journalism in *Digital Journalism*, vol. 3, issue 2, pp. 158-174
- Johnston, M. P. 2017. “Secondary Data Analysis: A Method of Which the Time has Come.” *Qualitative and Quantitative Methods in Libraries* 3 (3), pp. 619–626.

Joux, A. (ed.) (2023). *Journalisme et post-vérité*. Les Essentiels d'Hermès, Paris  
Kilby, A. 2018.  
“Provoking the Citizen: Re-examining the Role of TV Satire in the Trump Era.”  
*Journalism Studies* 19 (13), pp. 1934–1944.

Kulkarni S., Richard Thomas R., Komorowski M. & Lewis J. (2024), Innovating Online Journalism: New Ways of Storytelling, *Journalism Practice*, 17:9, pp. 1845-1863,

Lewis J., and S. Cushion. (2009) The Thirst To Be First: An Analysis of Breaking News Stories and Their Impact on the Quality of 24-Hour News Coverage in the UK. *Journalism Practice* 3(3). Pp. 304–318.

Lorusso, A., (2018), *Post-verità. Fra reality tv, social media e storytelling*, Laterza, Bari.

Niles J. D. (2010), *Homo Narrans: the Poetics and Anthropology of Oral Literature*, University of Pennsylvania Press, Philadelphia.

Norambuena, B. K., M. Horning, and T. Mitra. 2020. “Evaluating the Inverted Pyramid Structure Through Automatic 5W1H Extraction and Summarization.” Proc. of the 2020 Computation+ Journalism Symposium. Computation+ Journalism, 2020, 1–7.

McErlean K. (2018), Interactive Narratives and Transmedia storytelling, Routledge, New York.

McIntyre L. (2018), *Post-truth*, The Mit Press, Cambridge, (Ma).

Pavlik, J. 2000, The Impact of Technology on Journalism, *Journalism Studies* 1 (2): 229–237.

Pöttker, H. 2003. “News and its Communicative Quality: The Inverted Pyramid—When and Why Did It Appear?” *Journalism Studies* 4 (4): 501–511.

Rabe, R. 2007. “Inverted Pyramid.” In Encyclopedia of American Journalism, edited by S. L. Vaughn, 223–225. New York: Routledge.

Rizzuto F. (2023) “News consumption in Italy in the digital era. Trust, crisis and disinformation” in *Digitos. Revista de comunicación digital*, Univer- sitat de Valencia, vol. 9, Monográfico: *Consumo de noticias y evasión de noticias en el sistema de medios híbridos*, pp. 109-125.

Salmon C., (2008). *Storytelling. La fabbrica delle storie*, Fazi, Roma

Sorice M. (2020) Piattaformizzazione della sfera pubblica, *Comunicazione Politica* 3/2020, pp. 371-388, Il Mulino, Bologna.

Sorrentino C., Splendore S. (2022), *Le vie del giornalismo*, Il Mulino, Bologna.

Thomson, E. A., P. R. White, and P. Kitley. 2008. ““Objectivity” and “Hard News” Reporting Across

Cultures.” *Journalism Studies* 9 (2): 212–228.

Van Krieken, K. 2018. “Multimedia Storytelling in Journalism: Exploring Narrative Techniques in Snow Fall.” *Information* 9 (5): 123.

Van Krieken, K., and J. Sanders. 2019. “What is Narrative Journalism? A Systematic Review and an Empirical Agenda.” *Journalism*, 1464884919862056.

Wahl-Jorgensen, K. 2020. “An Emotional Turn in Journalism Studies?” *Digital Journalism* 8 (2): 175–194.

Walters, P. 2017. “Beyond the Inverted Pyramid: Teaching the Writing and all-Formats Coverage of Planned and Unplanned Breaking News.” *Teaching Journalism & Mass Communication* 7 (2): 9–22.

Ward, S. 2009. “Journalism Ethics.” In *The Handbook of Journalism Studies*, edited by K. Wahl-Jorgensen, and T. Hanitzsch, 295–309. New York: Routledge.

Van Dijck J., Poell T., & De Waal M. (2018). *The Platform Society: Public Values in a Connective World*, Oxford University Press (trad. it. *Platform society. Valori pubblici e società connessa*, Guerini e Associati, Milano 2023).

Yaros R. A. (2006), Is it the Medium or the Message? Structuring Complex News to Enhance Engagement and Situational Understanding by Nonexperts.” *Communication Research* 33 (4), pp. 285–309.

Zaganelli G., Marino T., Storytelling e Digital Journalism. Il caso della rivoluzione libica, *Comparatismi*, 1, 2016, 93-113 ISSN 2531-7547

## **Algoritmi editoriali. Una mappa per leggere l'automazione nelle newsroom – di Marzia Antenore, Dipartimento di Comunicazione e Ricerca Sociale Sapienza – Università di Roma**

*L'intelligenza artificiale entra nelle redazioni come filtro, segretario e persino direttore, spostando l'equilibrio tra controllo umano e automazione. Dai roomba che moderano i commenti agli editor-in-chief algoritmici, cambia il significato stesso di decisione editoriale*

“L'intelligenza artificiale invade le redazioni, ma è un giornalista o uno strumento?” si domandano Benjamin Mullin e Katie Robertson in un recente articolo del *New York Times*. Il quesito è interessante, ma fuorviante e parziale. È fuorviante perché invita a tematizzare l'Intelligenza Artificiale secondo una logica di opposizione, uomo-contro-macchina, che accentua la dimensione del conflitto con l'umano, una posizione oggi pragmaticamente insostenibile. Parziale, perché ignora che i sistemi intelligenti non aspirano soltanto a svolgere il lavoro giornalistico, ma a ridefinire le gerarchie interne all'informazione: si candidano a essere direttori responsabili, analisti editoriali, consulenti di marketing, attori cioè di una catena decisionale che va oltre la mera produzione di contenuti. Per alcuni questo scenario è spaventoso, l'adattamento impraticabile o comunque non auspicabile. Eppure, ridotta ai minimi termini, la storia del giornalismo è soprattutto una storia di adattamenti. Ogni rivoluzione tecnologica - dal telefono alla fotografia, dalla riproduzione meccanica all'informatizzazione - ha ridisegnato ruoli, pratiche e flussi di lavoro dentro le redazioni. L'intelligenza artificiale (AI), in questa traiettoria, non rappresenta una rottura ma una continuità. Negli ultimi anni alla macchina è stata trasferita una crescente autorità decisionale. Proprio come le macchine e la meccanizzazione hanno trasformato la produzione di oggetti materiali nel diciannovesimo e ventesimo secolo, l'informatica sta ora trasformando il lavoro intellettuale e cognitivo scaricandoli sui computer. La transizione alla “Seconda era delle macchine” implica una versione di automazione completa in cui un sistema funziona senza intervento umano, nonostante il lavoro umano di progettazione e manutenzione richiesto da tutti i sistemi progettati. Questa transizione, inevitabilmente, riguarda anche il lavoro delle newsrooms in cui si assiste a una progressiva delega all'Intelligenza Artificiale di molte fasi del newsmaking. Tuttavia, mai come oggi l'equilibrio fra controllo umano e automazione è diventato un campo di negoziazione strategica, in cui si giocano non solo le forme del lavoro giornalistico ma la stessa idea di controllo editoriale. L'ingresso nelle redazioni di sistemi di scrittura automatica basati sui *Large Language Models* ha solo reso più urgenti questioni cruciali. Dove tracciare il confine tra ciò che può essere delegato e ciò che deve restare prerogativa del redattore? Quanta agency possiamo trasferire agli algoritmi editoriali? Fino a che punto è lecito spingere l'automazione in campo giornalistico? In questo articolo discuto alcuni casi di automazione tenendo conto di due dimensioni interpretative: da una lato quanto autonomia viene ceduta all'algoritmo e, dall'altro, per quali ruoli redazionali e con quali conseguenza sul newsmaking. Comincerò dagli algoritmi-roomba.

### **Algoritmi-roomba**

Il Roomba è un aspirapolvere domestico prodotto da una azienda statunitense attiva anche nella robotica militare. Si tratta di un apparecchio di forma circolare che opportunamente programmato si stacca dalla stazione di ricarica e pulisce la stanza in completa autonomia per poi ritornare alla base. La descrizione del prodotto fornita dalla casa costruttrice è molto invitante:

“Il robot impara, e continua a imparare, dai ritmi della tua casa. Dall’analisi delle tue preferenze alla sincronizzazione con gli altri dispositivi smart, i nostri prodotti portano nuovi livelli di intelligenza che si adattano perfettamente alla tua vita così puoi dimenticare la pulizia dei pavimenti e concentrarti sulle cose che per te contano davvero”.

In sostanza, il Roomba è progettato in modo da minimizzare l’intervento umano per ridurlo a poche e semplici istruzioni iniziali, per il resto è perfettamente in grado di cavarsela da solo. Nel momento in cui il Roomba entra negli spazi domestici gli viene trasferita una autorità decisionale integrale: in base ai parametri impostati e alle sue capacità di apprendimento può fare le pulizie dove meglio crede. Chi non vorrebbe affidare compiti di routine e manovalanza a un sistema intelligente e collaborativo per dedicarsi ad attività più creative?

Da molti anni le redazioni giornalistiche si sono dotate di algoritmi editoriali che mi piace definire roomba, come i loro fratelli maggiordomi digitali. Non fanno le pulizie ma in un certo senso fanno il lavoro sporco velocemente e su larga scala in modo che anche i redattori possano occuparsi di “cose che contano davvero”.

Un noto roomba viene impiegato dal quotidiano *El País*, nella sua versione online. Stanca di occuparsi dell’hate speech, ossia dei commenti carichi di odio postati sotto agli articoli, la redazione del giornale spagnolo già da qualche anno ha preso una decisione radicale: la moderazione dei post è stata delegata a un algoritmo che, presumibilmente in base a una lista di parole chiave, filtra i commenti considerati inappropriati.

Anche il *New York Times* impiega un sistema di filtraggio automatico basato su un classificatore appreso dalla macchina per moderare i commenti sul suo sito. Utilizzando i dati su quali post sono stati contrassegnati da un moderatore come "tossici", l'algoritmo impara a classificare i commenti futuri come "tossici" o "non tossici" e, quando necessario, invita a riformulare la frase o ne impedisce la pubblicazione.

L’analoga con i compiti domestici non deve far pensare che questi algoritmi siano meno sofisticati o di poco conto. Tutt’altro. Come l’AI inscritta nello scheletro dell’apparecchio domestico prendono decisioni basandosi su valutazioni molto complesse. Ciò che fa di un algoritmo un Roomba riguarda solo l’autonomia con cui svolge il compito assegnato, autonomia intesa come l’estremo in una gamma di opzioni che combinano esseri umani e computer. Al livello più elevato – quello occupato da questi sistemi intelligenti - assistiamo a un trasferimento integrale di autorità decisionale in cui la macchina prende decisioni ignorando l’essere umano. A livello più basso di autonomia il computer offre all’umano un set completo di possibili alternative tra le quali scegliere. Nei roomba di *El País* e del *New York Times* il trasferimento di agency dall’uomo alla macchina è completo. Per questo, anche se svolgono un compito all’apparenza umile, i roomba possono rivelarsi piuttosto ambiziosi. L’agenzia di stampa britannica Reuters, per esempio, ha sperimentato processi di automazione molto estesi affidando a un roomba l’intera filiera del newsmaking: dalla individuazione della notizia, alla classificazione in temi sino alla valutazione sulla attendibilità e notiziabilità dell’evento. E’ così che nel 2019 vede la luce NewsTracer, una applicazione sviluppata in-house da Reuters. NewsTracer è un cacciatore di notizie che monitora Twitter in tempo reale sfruttando la tendenza umana a condividere

eventi che suscitano forti emozioni (una incredibile applicazione di *reverse engineering!*). Ogni giorno analizza fino a 12 milioni di tweet basandosi su 13 algoritmi di machine learning che filtrano il flusso dei messaggi, foto e video pubblicati online alla ricerca di focolai informativi ossia aree in cui improvvisamente aumenta la densità di parole chiave come *terrorismo*, *sparatoria*, *alluvione*. Rintracciati i focolai compie altre operazioni chiave: anzitutto valuta se l'evento è realmente accaduto oppure se si tratta di un falso allarme o *misinformation*, lo clusterizza assegnandolo a un genere (crime, terrorismo, evento naturale, politica, ecc.), e ne stima la notiziabilità intesa come capacità di stimolare engagement tra gli utenti della piattaforma. NewsTracer opera con due bussole principali che lo orientano nell'eliminare il rumore dalla rete: una lista di parole chiave, stilata da redazioni Reuters distribuite fra Stati Uniti, Europa, Asia e Africa; e una rete di circa cinquemila account di fonti attendibili, che include agenzie di stampa concorrenti, giornali, televisioni, istituzioni e forze dell'ordine. Quando individua un possibile evento, il sistema cerca di localizzarlo geograficamente e ne valuta l'attendibilità risalendo al tweet originale, verificando la reputazione delle fonti e analizzando la semantica dei commenti. Non si limita quindi a filtrare contenuti, ma interpreta, prioritizza e organizza le informazioni modellando la stessa idea di news in base a parametri computazionali.

Come mette in luce questa esperienza, prima di prendere una decisione i roomba compiono almeno quattro valutazioni molto sofisticate per una macchina: assegnano una priorità (prioritizing), classificano (classifying), associano (associating) e filtrano (filtering). Spesso, inoltre, queste decisioni vengono poi composte in attività informative di livello superiore.

Vale la pena soffermarsi su queste valutazioni perché è *la loro qualità* a determinare quanto possiamo spingere l'automazione in campo giornalistico.

La prima valutazione, **dare priorità**, consiste nel decidere quali elementi mostrare prima o attribuire maggiore visibilità. Gli algoritmi dei social network, ad esempio, determinano l'ordine dei post nei feed in base a criteri di interazione o rilevanza stimata, mentre i motori di ricerca stabiliscono la posizione dei risultati sulla pagina in base alla pertinenza, all'affidabilità e all'aggiornamento delle fonti. Anche nelle redazioni giornalistiche che adottano strumenti di raccomandazione automatica, la priorità può essere assegnata a contenuti "di tendenza" o ritenuti di maggiore interesse pubblico.

La seconda operazione, **classificare**, riguarda l'attribuzione di etichette o categorie ai contenuti. Gli algoritmi di riconoscimento testuale, per esempio, distinguono tra articoli di politica, sport o cultura; i sistemi di filtraggio delle email separano i messaggi legittimi dallo spam; e i modelli di analisi del linguaggio naturale usati nelle redazioni possono clusterizzare i testi per tema, tono o fonte. La classificazione è quindi un passaggio necessario per organizzare l'informazione e renderla gestibile in modo automatico.

La terza funzione, **associare**, permette di collegare elementi diversi individuando somiglianze o relazioni. È la logica che guida le raccomandazioni di piattaforme come Amazon ("prodotti spesso acquistati insieme") o Spotify ("brani simili"). In ambito giornalistico, la stessa logica può suggerire "articoli correlati" a quello appena letto, favorendo percorsi di approfondimento che rispondono alle preferenze dell'utente.

Infine, **filtrare** significa selezionare ciò che viene mostrato e scartare ciò che viene escluso. I filtri possono operare per motivi di sicurezza (come nel caso dei contenuti violenti o offensivi rimossi da El País e dal New York Times) oppure per finalità informative, ad esempio limitando i risultati di ricerca a una certa area geografica o a una lingua specifica. In tutti i casi, il filtraggio modella in modo decisivo l'esperienza informativa, influenzando ciò che l'utente percepisce come rilevante o accessibile.

Queste quattro operazioni di rado agiscono separatamente: nella maggior parte dei sistemi, **si combinano in modo dinamico**. Un algoritmo di raccomandazione giornalistica ad esempio può classificare gli articoli per tema, filtrare quelli datati o ridondanti, associare storie simili e, infine, dare priorità a ciò che ritiene più interessante per l'utente.

Come abbiamo anticipato, la qualità di queste valutazioni determina fino a che punto è possibile spingere l'automazione. Nel caso di una agenzia di stampa come Reuters che fonda la sua attività sulle breaking news l'automazione si dimostra basata su valutazioni di qualità elevata per i fini di *news detection* che si era proposta. Uno studio comparativo sulla velocità con cui Reuters scava le notizie, soprattutto quelle inattese, mette in rilievo come l'automazione ha consentito a Reuters di individuare le breaking news entro 8 minuti dal verificarsi dell'evento, contro i 40 minuti di altre agenzie di stampa e gli 80 minuti dei global media.

### Segretari algoritmici

Non bisogna confondere i roomba con gli algoritmi-segretario, di cui comunque sono parenti stretti. Sebbene le valutazioni da compiere prima di prendere una decisione siano le medesime, un assistente intelligente si muove in funzione di servizio e dipendenza dall'essere umano a cui propone alternative valide e risorse sulla base delle quali assumere decisioni. Il segretario algoritmico ha un livello di autonomia inferiore rispetto a un roomba, così come un segretario umano prepara i documenti sui quali il dirigente appone la firma. Sistemi come *Wordsmith*, adottato dall'*Associated Press* già dal 2015, incarnano questo paradigma ibrido tra umano e macchina: generano automaticamente articoli basati su dati - per esempio risultati sportivi o rapporti finanziari - ma lasciano al redattore il compito di revisionare, contestualizzare e firmare i testi. Se lo stile o il tono dell'articolo generato sono poco convincenti il redattore può chiedere alla macchina di riscriverlo solo facendo click su un pulsante. Potenzialmente le riformulazioni dello stesso pezzo sono infinite. Da sempre all'avanguardia nell'automazione, nel 2025, Associated Press ha usato segretari algoritmici per esaminare rapidamente decine di migliaia di pagine di documenti relativi agli assassinii del presidente John F. Kennedy, del reverendo Martin Luther King Jr. e di Robert F. Kennedy. Gli strumenti hanno reso i documenti ricercabili e li hanno riassunti, consentendo probabilmente ai giornalisti di risparmiare giorni di lavoro.

Sistemi intelligenti come *Echobox* o *True Anthem* adottati anche da alcune redazioni italiane sono collaboratori altrettanto efficienti: analizzano modelli di clic, cronologie di engagement e performance passate per prevedere non solo quando pubblicare, ma anche che cosa pubblicare su canali social in base a quale contenuto ottiene più attenzione tra due simili (una pratica nota come A/B test). Echobox dichiara che i propri clienti registrano un aumento medio del 57% del traffico Facebook e del 100% su Twitter quando lasciano che sia l'algoritmo a gestire il momento della pubblicazione. Non sorprende. Uno dei vantaggi principali dell'automazione è la capacità adattiva: l'algoritmo si aggiorna costantemente in risposta ai cambiamenti degli algoritmi di distribuzione delle piattaforme. In questo modo, i sistemi editoriali rispondono agli algoritmi con altri algoritmi, in una forma di ottimizzazione reciproca: un algoritmo impara a ottimizzare rispetto all'altro, eliminando la necessità di intuire manualmente quali elementi del contenuto sono premiati dalla piattaforma. Al redattore non resta che affidarsi alla macchina e "apporre la firma" abilitando la pubblicazione. Anche in questi casi sembra che l'algoritmo sia in grado di prendere decisioni di qualità, se misuriamo la qualità come aderenza con il risultato atteso e auspicabile.

Se ci fosse un mantra di questi sistemi intelligenti sarebbe “Automate what computers do best, let people do the rest.” Automatizza ciò che le macchine sanno fare meglio e lascia il resto agli esseri umani.

Ma cosa succede quando le macchine prendono decisioni alla stregua (o migliori) dell’essere umano per compiti e ruoli apicali?

I casi che abbiamo visto non riguardano questo tipo di mansioni: la moderazione dei commenti, la scelta di quando pubblicare un post restano nell’ambito delle decisioni operative più che in quelle strategiche, anche nel caso di una automazione spinta come quella di Reuters dove l’applicazione di news detection si comporta alla stregua di un redattore (un ottimo giornalista investigativo!).

Dobbiamo quindi introdurre una seconda dimensione, ossia quella del livello editoriale, lungo la quale si realizza l’attribuzione della delega alla macchina. Se siamo abituati a pensare all’algoritmo come a un circuito ibrido dove la macchina ha il dominio sulla velocità di esecuzione e la persona sul giudizio, nei casi che presentiamo adesso vediamo come l’AI si insedia in ambiti in cui il giudizio è di norma attribuito all’essere umano. Il primo di questi casi riguarda l’ambito della supervisione editoriale.

### AI Editor-in-Chief

Il Globe and Mail, storico quotidiano canadese con oltre 170 anni di storia, utilizza una intelligenza artificiale chiamata Sophi per la regolazione del flusso informativo: circa il 99% degli articoli che compaiono in homepage viene scelto in modo automatico, mentre la redazione umana seleziona solo le tre o quattro notizie principali del momento.

Dal punto di vista computazionale, questo sofisticato algoritmo di filtraggio dei contenuti usa tecniche di Elaborazione del Linguaggio Naturale (NLP) per estrarre automaticamente gli argomenti principali di ogni articolo, classificarli, e suggerire ai lettori registrati quali *topics* seguire sfruttando set di metadati per migliorare la raccomandazione dei contenuti. Ma la cosa più interessante è che Sophi ha uno strano modo di interpretare la *qualità* dell’informazione. Ogni pochi minuti monitora il web per verificare quanto un articolo venga condiviso e discusso, quindi lo valuta unicamente in base alla sua capacità di attrarre nuovi abbonati. In sostanza Sophi traduce criteri economici e comportamentali in valutazioni editoriali. Non solo dunque il potere decisionale, prerogativa dei caporedattori, qui si sposta verso la macchina, ma siamo anche di fronte a una macchina addestrata a valutare il valore di una notizia non più in termini di rilevanza pubblica ma di potenziale di abbonamento. La call to action a cui si ispira il filtraggio non è dunque click, like o share, ma pay, ossia un’azione che pochi sono disposti a fare online. Va da sé che dietro il sipario ci sono sempre persone: designer, redattori, giornalisti, data e computer scientist che contribuiscono in modo diretto o indiretto. Chi progetta o supervisiona gli algoritmi prende decisioni editoriali chiave: come parametrizzare il sistema, quali impostazioni predefinite adottare, a quali segnali prestare attenzione e, di fatto, quali valori incorporare nel nucleo del sistema stesso. Ma una volta prese queste decisioni a monte, l’intera autorità decisionale viene trasferita a valle della filiera. Nell’esperienza di Sophi la selezione di articoli da piazzare in homepage riflette la conoscenza giornalistica e le aspettative di genere, stile, tono e lessico più appropriati per i lettori potenziali. La linea editoriale dell’organizzazione è intrecciata nella rete neurale artificiale attraverso dati, regole e basi di conoscenza. Così come le persone e i loro valori sono incorporati nell’intero ecosistema di algoritmi umani che costituiscono i media. Il caso di Sophi

segnalà così un passaggio simbolico prima che tecnologico: l'intelligenza artificiale non è più un assistente o un filtro, ma un vero e proprio Editor-in-Chief.

### L'algoritmo diventa direttore marketing

In Svezia, il gruppo MittMedia - principale editore di informazione locale con 28 testate e una storia che risale a oltre due secoli, per un pubblico complessivo di 1,8 milioni di lettori al giorno - rappresenta l'altra faccia di questa delega algoritmica: non più l'AI editor-in-chief, ma l'AI direttore del marketing.

L'intelligenza artificiale di MittMedia sceglie i contenuti da mettere in home page e propone a ciascun utente un'informazione personalizzata in base alle sue preferenze di lettura. L'algoritmo crea un identikit del lettore combinando dati sociodemografici e comportamentali e, su questa base, clusterizza gli utenti in gruppi di interesse. L'obiettivo non è solo aumentare le visite, ma prevenire le disdette: la fedeltà quotidiana alla piattaforma si traduce in una maggiore fedeltà all'abbonamento.

Per farlo, MittMedia utilizza uno strumento di analisi visiva in cui le bolle del grafico rappresentano gli articoli pubblicati durante la giornata, suddivisi per tema (incidenti, cronaca, affari locali, ecc.). L'algoritmo evidenzia gli argomenti che "lievitano" in termini di performance, mostrando -ad esempio - che le notizie sugli affari locali generano il maggior numero di abbonamenti. A quel punto, i manager chiedono ai direttori delle testate di produrre più contenuti su quel tema, indicando anche forma e taglio attraverso linee guida interne.

L'esperienza di MittMedia mostra quello che è stato definito "effetto salmone dell'automazione": un movimento che, partendo dalla fase finale del processo produttivo - l'ottimizzazione dei post per i social e l'audience digitale - risale progressivamente la corrente fino a rimodellare la cultura redazionale nel suo complesso. Così, l'automazione non si limita più a gestire la distribuzione, ma comincia a influenzare la produzione stessa delle notizie, fino a modificare il play-out delle pagine cartacee e le priorità editoriali tradizionali.

A differenza di quanto avviene nella redazione dei colleghi canadesi, la direzione marketing usa questi sistemi intelligenti solo come strumento di consulenza per pianificare le scelte strategiche di marketing, mentre l'agency vera e propria rimane sotto il controllo umano.

La figura 1. Propone dunque una mappa dell'automazione delle newsrooms che fotografa un ecosistema composito, dove l'intelligenza artificiale non è solo uno strumento ma un attore redazionale a tutti gli effetti con differenti livelli di autonomia e diversi ruoli nel processo di newsmaking.

Nella parte inferiore della matrice, gli algoritmi operano come **assistenti o segretari digitali**: automatizzano compiti di routine - dalla generazione di articoli basati su dati (Wordsmith), alla sintesi di documenti di migliaia di pagine (Associated Press), alla moderazione dei commenti (El País, NYT) - liberando tempo e risorse per i redattori umani. In questi casi, il controllo resta saldamente umano e la macchina agisce per supporto, non per sostituzione.

Man mano che ci si sposta verso l'alto, l'autonomia cresce e il ruolo dell'AI si fa più strategico. Sistemi come quelli adottati da **MittMedia** e Globe and Mail incarnano una nuova frontiera, dove gli algoritmi assumono funzioni apicali: definiscono priorità, gerarchie e persino criteri di valore economico o editoriale. Se nel primo caso l'AI agisce come un consulente di marketing, nel secondo diventa di fatto

un *Editor-in-Chief* computazionale, capace di plasmare la visibilità delle notizie sulla base della “payworthiness”.

Nel complesso, la mappa rivela una **verticalizzazione dell’automazione**: dagli strumenti che assistono, agli algoritmi che decidono. E mostra che il punto critico non è per quale ruolo possiamo o non possiamo cedere il comando, ma fino a che punto la macchina è in grado di fare valutazioni qualitativamente soddisfacenti.

**Figura 1. Una mappa dell’automazione delle newsrooms lungo le due dimensioni di autonomia e livello decisionale.**

	<b>Bassa autonomia decisionale</b> (Controllo umano)	<b>Alta autonomia decisionale</b> (Autonomia algoritmica)
<b>Figura apicale</b> <i>(AI come direzione editoriale)</i>	MittMedia (Svezia) – AI Direttore Marketing: individua temi più redditizi, orienta la produzione. AI advisor: supporta le decisioni ma non agisce.	Sophi (Globe and Mail) – AI Editor-in-Chief: decide il 99% della homepage in base alla payworthiness.
<b>Figura operativa</b> <i>(AI come assistente o segretario)</i>	Wordsmith (Associated Press) – genera articoli da dati, redattore revisiona e firma. Echobox / True Anthem – ottimizzazione dei contenuti social, suggerisce ma non decide.	NewsTracer (Reuters) – scopre notizie su Twitter, individua “focolai” informativi e li classifica. El País / NYT Comment Moderation – filtra automaticamente i commenti tossici.

### Lo strano caso di ChtatGpt

L'inizio del 2019 segna uno spartiacque nel campo del giornalismo computazionale: un algoritmo diventa coautore di un pezzo giornalistico. Il 31 gennaio *The Guardian Australia* pubblica il primo articolo interamente generato da un sistema di AI, dedicato alle donazioni ai partiti politici da parte dei cittadini. Non siamo di fronte ad una lista di dati rastrellata dal web e inserita in un template, come nel caso di WordSmith, ma assistiamo a una forma embrionale di interpretazione dell'informazione generata automaticamente da ReporterMate, “an experimental automated news reporting system”. Nel 2020 OpenAI di Elon Musk lancia GPT-3: una rete neurale che usa strumenti di Natural Language Processing a cui è richiesto di cimentarsi con tema, con una organizzazione autonoma del ragionamento. Così una macchina scrive la sua apologia su *The Guardian* “Are you scared yet, human?” chiede beffarda al pubblico di lettori del quotidiano, inaugurando una fase in cui l'AI si appropria del gesto stesso di scrivere. Così, nel 2022, *The Economist* apre il numero di maggio con un editoriale sulla guerra in Ucraina scritto integralmente da un modello linguistico artificiale. Recentemente, Fortune e Business Insider, hanno valutato l'utilizzo dell'intelligenza artificiale per scrivere articoli completi; Newsquest, una catena di quotidiani britannica di proprietà di USA Today Company, impiega oltre 30 giornalisti che utilizzano l'intelligenza artificiale per approfondire le storie. È qui che gli algoritmi ambiscono a competere in un campo che fino ad allora sembrava precluso ad

una intelligenza sintetica ossia quello dello storytelling. A onore del vero, già nel 1958 i ricercatori dell'IBM descrissero un programma in grado di estrarre automaticamente un riassunto da un documento di ricerca o da un articolo di giornale. Se si trattava di un frammento rappresentativo, l'algoritmo lo estraeva e lo aggiungeva al riepilogo. Nel 2013, Yahoo! ha iniziato a utilizzare la tecnologia di riepilogo nella sua app di notizie per assemblare le informazioni di diversi articoli di notizie in un unico briefing. La tecnologia per analizzare il testo tramite computer esiste da decenni. Ma le valutazioni automatiche necessarie per riassumere un articolo hanno raggiunto solo di recente un livello di qualità che consente ai riassunti di avere un valore effettivo nel mercato dei media.

ChatGPT è emblematico di questa svolta. Il suo lavoro cognitivo comprende funzioni di elaborazione dati, ma soprattutto cattura l'idea di compiti di manipolazione analitica delle informazioni testuali tipicamente associati all'intelligenza umana. A differenza dei suoi antenati, ChatGPT è uno strumento dotato di una proattività radicale: non solo compone testi ma suggerisce i prossimi passaggi, anticipa le intenzioni dell'utente e, in un certo senso, lo addestra a un nuovo tipo di interazione.

Personalmente, credo che una delle sue affordances più insidiose sia la passivizzazione dell'interlocutore umano: la tentazione di accettare ciò che propone senza riformularlo, fidandosi della sua fluidità è davvero irresistibile. E' quanto deve aver sperimentato l'incauto redattore di una testata italiana che ha trascurato di omettere l'invito rivoltogli da ChatGPT - che suonava più o meno come "Se vuoi posso trasformalo in un reportage da pubblicare su un magazine specializzato" - pubblicandolo accidentalmente sul quotidiano. Ma poiché nessuno è del tutto immune alla tentazione di fare alla svelta sono emersi simili errori imbarazzanti anche da importanti testate come Bloomberg, Business Insider e Wired.

Anche per evitare situazioni di questo tipo molte testate internazionali hanno iniziato a dotarsi di policy interne sull'uso dell'AI, spesso con un approccio prudenziale: vietare o limitare l'impiego dell'Intelligenza Artificiale nella stesura di articoli, nella raccolta di fonti o nella selezione delle immagini. Ma le policy - come ogni forma di regolazione - inseguono la realtà più che anticiparla: Axios, un sito di news con base ad Arlington, Virginia, sta già automatizzando i riassunti delle notizie utilizzando ChatGPT per selezionare le news più rilevanti del giorno. A fronte di queste esperienze, non sembra ancora chiaro quale ruolo possa ricoprire l'AI generativa all'interno delle redazioni, se apicale o di servizio.

ChatGPT, al momento, rimane uno strano caso occupando una posizione mobile nella mappa dell'automazione delle newsrooms tracciata nella figura 1. Nella maggior parte dei contesti redazionali - dove è utilizzato per scrivere bozze, riassunti o traduzioni sotto stretto controllo umano - si colloca nel quadrante in basso a sinistra, quello dell'assistente cognitivo o segretario algoritmico: un supporto operativo, privo di autonomia decisionale.

Quando, tuttavia, viene impiegato per **generare testi completi senza revisione** - come nel caso del giornalista che ha pubblicato inavvertitamente la risposta integrale del modello (ma esistono anche casi virtuosi come quello di Axios) ChatGPT si sposta verso il **quadrante inferiore destro**, assumendo un ruolo più autonomo ma ancora subordinato: una forma di *ghostwriter artificiale*.

Rimane aperta la questione se, ed eventualmente in che forma, un'intelligenza generativa possa occupare posizioni apicali all'interno delle redazioni - una possibilità che mi sembra intrecciata con quella del grado di libertà di parola attribuibile ai bot sulle piattaforme.

Alcune recenti sperimentazioni suggeriscono che questo scenario non è più del tutto ipotetico.

Emblematico è il caso di IDEIA – **Intelligent Engine for Editorial Ideation and Assistance** – progetto sviluppato in collaborazione con il gruppo mediatico brasiliano *Sistema Jornal do Commercio de Comunicação (SJCC)*. Si tratta di un sistema basato su modelli linguistici generativi e progettato per **automatizzare la fase di ideazione editoriale**, combinando dati di tendenza in tempo reale (attraverso la *Google Trends API*) con la generazione di titoli e sommari contestuali (*Google Gemini API*). IDEIA, dunque, non si limita a supportare la scrittura o a riassumere contenuti esistenti: **interviene nella fase preliminare di definizione dell'agenda**, proponendo temi, angolature e linee di sviluppo, sulla base di trend emergenti e analisi predittive.

Questi esempi mostrano come la traiettoria delle intelligenze generative stia progressivamente spostandosi verso l'alto della mappa, avvicinandosi al quadrante apicale: dagli algoritmi assistivi e subordinati all'intervento umano, alle intelligenze direttive, in grado di incidere sul coordinamento e sulla strategia editoriale. In scenari sperimentali come questi - dove gli output dei modelli vengono usati per orientare scelte editoriali, titolazioni o gerarchie di notizie - ChatGPT e i suoi simili tendono a risalire la matrice, avvicinandosi al ruolo di co-decisorì redazionali, ovvero a quella soglia critica in cui l'automazione si confonde con la direzione.

La parabola del giornalismo, letta alla luce di questa mappa dell'automazione, rivela che il confine fra controllo umano e delega algoritmica non è mai netto, ma costantemente negoziato. I roomba, i segretari e gli editor-in-chief digitali rappresentano gradi diversi di questa delega, ma anche fasi di un unico processo evolutivo in cui le macchine apprendono, sperimentano e si spostano progressivamente verso l'alto della gerarchia decisionale. Ciò che cambia non è soltanto il modo in cui le redazioni producono le notizie, ma il significato stesso di "decisione editoriale". Se fino a pochi anni fa essa implicava un atto deliberativo, un giudizio situato e argomentato, oggi tende a dissolversi in un calcolo probabilistico, ottimizzato per la performance. I Large Language Models come ChatGPT ne sono la frontiera più ambigua: strumenti di assistenza, ma anche potenziali coautori, editor e interpreti della realtà. Il futuro delle redazioni dipenderà dalla capacità di trasformare questa ambiguità in governance: progettare sistemi in cui l'AI non solo esegua, ma rispecchi valori giornalistici esplicativi - veridicità, responsabilità, pluralismo - inscritti nei suoi parametri. Come abbiamo ribadito è la qualità delle valutazioni umane inscritte nel DNA della macchina a determinare fino a che punto è possibile spingere l'automazione.

## Bibliografia

Antenore, M., Bruno, M., & Valentini, E. (2025, settembre 18–20). *Strategie e spazi di autonomia delle testate giornalistiche alla prova dell'Intelligenza Artificiale: il caso italiano nel contesto europeo*. Intervento presentato al convegno Associazione Italiana di Sociologia (AIS), Bergamo.

THE ASSOCIATED PRESS. AP, *NCAA to grow college sports coverage with automated game stories*. In: Associated Press, 4 marzo 2015. Disponibile su: <https://apnews.com/article/automated-insights-ncaa-college-sports> [Consultato: 11 novembre 2025].

DELGADO, Pablo. *How El País used AI to make their comments section less toxic*. In: Google News Initiative, 21 marzo 2019. Disponibile su: <https://newsinitiative.withgoogle.com/articles/how-el-pais-used-ai-to-make-their-comments-section-less-toxic/> [Consultato: 11 novembre 2025].

Diakopoulos, N. (2019). *Automating the News: How Algorithms Are Rewriting the Media*. Cambridge, MA: Harvard University Press.

ECHObox. *AI for social media publishing*. In: *Echobox Official Website*, 2023. Disponibile su: <https://www.echobox.com/> [Consultato: 11 novembre 2025].

Knaus, C. (2019, January 31). *Political donations plunge to \$16.7m – down from average \$25m a year*. The Guardian. Retrieved from <https://www.theguardian.com/australia-news/2019/feb/01/political-donations-plunge-to-167m-down-from-average-25m-a-year> [Consultato: 11 novembre 2025].

Fontanarosa, A. (2020). *Giornalisti robot: L'intelligenza artificiale in redazione. Prove tecniche di news revolution*. UlisseAspettaPenelope.

**Liu, X., Nourbakhsh, A., Li, Q., Fang, R., & Shah, S. (2017).** *Reuters Tracer: A large-scale system of detecting and verifying real-time news events from Twitter*. Proceedings of the 25th ACM International Conference on Information and Knowledge Management (CIKM '17), 207–216. Association for Computing Machinery. <https://doi.org/10.1145/3132847.3132880>.

**Mezza, M. (2018).** *Algoritmi di libertà: La potenza del calcolo tra dominio e conflitto*. Donzelli Editore.

Mullin, B., & Robertson, K. (2025, November 7). *AI Sweeps Through Newsrooms, but is it a journalist or a tool?* The New York Times. Retrieved from <https://www.nytimes.com/2025/11/07/business/media/ai-news-media.html?smid=nytcore-android-share>

**Santos, V. B. et al. (2025), IDEIA: A Generative AI-Based System for Real-Time Editorial Ideation in Digital Journalism**, arXiv preprint arXiv:2506.07278v1.

<https://arxiv.org/abs/2506.07278v1>

**SOPHI.IO / THE GLOBE AND MAIL.** *How The Globe and Mail uses Sophi automation and analytics to power journalism*. In: *Sophi.io Case Studies*, 2020. Disponibile su: <https://www.sophi.io/> [Consultato: 11 novembre 2025].

The Guardian. *A robot wrote this entire article. Are you scared yet, human?* In: *The Guardian*, 8 settembre 2020. Disponibile su: <https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3> [Consultato: 11 novembre 2025].

## **Bulimia social: quando l'iperconnessione diventa malattia – di Marino D'Amore, Docente di Sociologia generale presso Università degli Studi Niccolò Cusano**

*L'iperconnessione genera una forma di bulimia digitale dove esistere significa essere visibili online. Gli algoritmi amplificano questo bisogno compulsivo di attenzione, trasformando gli utenti in produttori instancabili di contenuti per ottenere gratificazione virtuale*

L'esplosione dei social network ha modificato radicalmente la socializzazione, le dinamiche relazionali e la percezione identitaria. L'individuo always connected vive in uno spazio sociocomunicativo coinvolgente e pervasivo, in cui l'interazione è continua e il silenzio, l'assenza, la disconnessione, sono percepiti come vuoti da colmare, come un pericolo incombente. In tale scenario prende forma una nuova patologia sociale e simbolica: la bulimia social. Con questo termine si intende un comportamento ossessivo-compulsivo, caratterizzato da una feroce sovraesposizione digitale, da consumo di contenuti e da relazioni virtuali, comportamenti accompagnati dal bisogno di gratificazione che si attualizza attraverso like, visualizzazioni e commenti. Insomma, una vera e propria modalità di esistenza e autorappresentazione, che si fonda sulla performatività del sé, sull'iperconnessione e la dipendenza (Bauman, 2003). Questa sindrome sociale rappresenta una delle esteriorizzazioni più evidenti delle dinamiche postmoderne: l'individualismo narcisistico, la spettacolarizzazione e l'erosione irreversibile del confine tra pubblico e privato.

### **La bulimia social**

La bulimia social si caratterizza per un' iperproduzione e consumo, direttamente proporzionale, di contenuti digitali, dovuta a una necessità costante di attenzione e di presenza nell'arena pubblica virtuale. Il concetto riprende analogicamente quello della bulimia alimentare, dove a fasi di ingurgitamento compulsivo seguono sensi di colpa, ansia e il bisogno di purificazione.

Nel contesto digitale, l'*overposting*, l'implementazione compulsiva di stories, reels, aggiornamenti e selfie, non rappresentano soltanto pratiche comunicative, ma espressioni sintomatologiche di un disagio profondo. L'utente non comunica per condividere, ma per esistere. La sua identità si fonda sulla visibilità e sull'interazione continua: “più mi vedono, più sono” (Marwick, 2013).

Le piattaforme social diventano il luogo in cui si consuma una relazione tra sé e gli altri, incentrata sulla prestazione e sull'autorappresentazione effimera. Il profilo personale diventa il palco dell'identità e la relazione con gli altri è mediata dal bisogno di essere riconosciuti, approvati, cliccati.

In un recente discorso in Piazza San Pietro, Papa Leone XIV ha affrontato in modo diretto e coraggioso il tema della “bulimia social”, definendola una malattia dell'anima che svuota il cuore mentre promette di riempirlo di consensi. Il Pontefice ha sottolineato come il bisogno incessante di essere presenti, di mostrarsi continuamente, non è un esercizio di libertà ma una nuova forma di schiavitù che logora la verità del sé e l'autenticità dei legami. Papa Leone XIV ha spesso richiamato l'urgenza di una ecologia dell'interiorità, capace di resistere all'overload comunicativo e all'ossessione per la visibilità. Il Pontefice ha parlato della necessità di riscoprire il valore del silenzio, della discrezione, che non cerca i riflettori ma costruisce relazioni vere e profonde. Tale riflessione si inserisce in una linea di pensiero già sviluppata da Papa Francesco, il quale aveva più volte criticato la cultura

dell'apparenza superficiale promossa dai media digitali (Francesco, 2020). Tuttavia, Papa Leone XIV porta questa critica a un livello ancora più strutturale, parlando apertamente di anoressia relazionale e bulimia dell'io, per descrivere il paradosso per cui l'iperesposizione del sé in rete coincide spesso con un vuoto identitario e affettivo. Il contributo del Papa non è solo religioso o morale, ma anche profondamente sociologico, poiché evidenzia le dinamiche di potere, alienazione e consumo che regolano la vita online. In tal senso, il suo intervento rappresenta un forte appello a ripensare l'etica della comunicazione digitale, mettendo al centro la persona nella sua interezza e non nella duplice veste di prodotto/contenuto.

### **Cause socioculturali e effetti sulle relazioni sociali della bulimia social**

L'origine di questo comportamento affonda le radici in vari fattori culturali e strutturali. Innanzitutto, la società della performance e del consumo (Han, 2012) ha creato un individuo che deve costantemente brandizzare sé stesso. Il soggetto contemporaneo è imprenditore della propria immagine e il social network è il luogo privilegiato in cui si concretizza questo lavoro simbolico.

In secondo luogo, la perdita di riferimenti comunitari stabili e delle old agency, come famiglia, scuola, gruppo religioso, ha aumentato la necessità di appartenere a nuove comunità simboliche, spesso fondate sulla condivisione digitale. Il senso di appartenenza si costruisce attraverso la partecipazione online.

Infine, la cultura dello spettacolo (Debord, 1967) e la logica della mediatizzazione della vita quotidiana (Couldry, 2012) contribuiscono a creare un contesto in cui non esistere nel flusso mediale significa, in un certo senso, non esistere affatto. La bulimia social è figlia diretta di questa condizione.

I principali effetti della bulimia social si registrano a livello psicologico e relazionale. Dal punto di vista psichico, si osservano sintomi legati all'ansia da prestazione, all'insicurezza, alla dipendenza dalla visibilità: l'assenza di notifiche può generare frustrazione, senso di fallimento e stati depressivi (Twenge, 2017). Le relazioni vengono spesso ridotte a interazioni superficiali, a legami deboli (Granovetter, 1973), che non riescono a offrire un vero supporto emotivo. Sul piano sociale, la bulimia digitalizza i rapporti, riducendo la complessità delle relazioni a metriche (cuori, like, reaction). L'alterità non si presenta più attraverso l'unicità della persona ma come pubblico standardizzato e invisibile. In questo scenario, si parla sempre più spesso di "solitudine connessa", una forma nuova di isolamento relazionale (Turkle, 2011).

### **Narcisismo digitale e desiderio di visibilità**

Uno dei motori principali della bulimia social è il narcisismo digitale. Le piattaforme incoraggiano la costruzione di un'immagine idealizzata di sé, costantemente curata, editata, filtrata. Il concetto di "panottico orizzontale" (Foucault, 1975) diventa fortemente attuale: ciascuno osserva e viene osservato, in un circuito infinito di visibilità reciproca. L'identità non è più un dato, una caratteristica, ma una narrazione continua. Chi non aggiorna il proprio profilo rischia di scomparire, come se l'identità fosse in stand by.

Il narcisismo digitale non è il risultato di uno stato patologico in sé, ma diventa tale quando è compulsivo, dipendente e insaziabile. Il soggetto si trasforma in una macchina produttiva di contenuti, soggetto a cicli alternati di euforia e delusione, e vittima del giudizio costante degli altri (Lasch, 2018). Non si può comprendere la bulimia social senza analizzare il funzionamento tecnico delle piattaforme. Gli algoritmi sono progettati per massimizzare il tempo di permanenza, la produzione e il consumo di contenuti. Il meccanismo di “reward” intermittente (Zuboff, 2019) produce un effetto simile a quello del gioco d’azzardo: non si sa mai quale contenuto avrà successo, ma si è sempre incentivati a produrne uno nuovo per raggiungere quello stesso successo.

La logica dell’engagement, che premia la visibilità a scapito della profondità, alimenta l’ansia performativa. Chi si ferma, perde; chi non si espone, si eclissa. Le piattaforme non sono neutrali, ma agiscono come amplificatori piramidali di bisogni e dinamiche di tipo compulsivo (Pariser, 2012).

### Nativi digitali e bulimia social

I nativi digitali sono la macrocategoria particolarmente esposta a questa dinamica, per ragioni evolutive e culturali. Durante l’adolescenza, il bisogno di appartenenza e di identità è massimo e i social network offrono un palcoscenico in cui sperimentare il proprio ruolo, ma anche un campo di prova, elargitore di giudizi e confronti. Il fenomeno del *compare and despair* è amplificato dalla continua esposizione a contenuti idealizzati e appositamente selezionati da dinamiche algoritmiche sempre più invasive. L’identità si costruisce sulla base del feedback e il valore personale è spesso misurato sui numeri e le statistiche generate da quest’ultimo. Questo porta a un’instabilità emotiva, alla paura dell’esclusione (FOMO) e, nei casi più gravi, a disturbi del comportamento, isolamento e depressione (Keles et al., 2019), sintomi depressivi, ansia sociale, disturbi del sonno (Andreassen, 2015). Il bisogno incessante di connessione compromette la capacità di concentrazione e l’equilibrio emotivo e il sovraccarico informativo-relazionale genera una forma di stress cognitivo che si traduce in alienazione (Carr, 2011).

### Conclusioni

La bulimia social rappresenta uno dei fenomeni più rilevanti e al contempo più sottovalutati del nostro tempo, poiché coinvolge la soggettività di ogni utente. Non si tratta solo di un disturbo comportamentale individuale, ma di un riflesso sistematico di una società interconnessa, iperperformativa e continuamente affamata di riconoscimento. In questo scenario, la distinzione tra benessere digitale e disagio psicosociale si fa sempre più labile. Il corpo sociale sembra aver interiorizzato una logica compulsiva che si regge su due assi fondamentali: la necessità di visibilità continua e la misurazione del valore personale attraverso parametri esterni, spesso riduttivi e meramente quantitativi (come follower, like, visualizzazioni). La conseguenza è la trasformazione dell’individuo in brand, impegnato in una narrazione permanente di sé che riduce la complessità dell’identità alla sua rappresentazione socialmente desiderabile.

A livello sociologico, possiamo parlare di una colonizzazione della sfera dell’intimità da parte dell’economia dell’attenzione (Citton, 2014), in cui anche le emozioni diventano un valore misurabile. La vita privata diventa contenuto, la vulnerabilità viene esibita come strategia di engagement e la verità rischia di diventare un simulacro calcolato. In questa logica, anche il dolore viene performato

per ottenere empatia virtuale e costituisce il paradosso di una società che, pur desiderando autenticità, la sostituisce con una “simulazione affettiva”, come ha osservato Sherry Turkle (2015).

La soluzione al problema non può che essere multidisciplinare e multifattoriale:

- *educativa*: con l’inserimento di programmi scolastici di percorsi alfabetizzazione digitale e affettiva che aiutino i giovani a distinguere tra identità reale e rappresentazione online, a sviluppare il pensiero critico e a costruire un rapporto sano con i media.
- *tecnologica*: con un maggiore impegno da parte delle piattaforme social nell’introdurre strumenti di moderazione del tempo d’uso, notifiche intelligenti, indicatori di benessere digitale, e modelli algoritmici meno aggressivi.
- *culturale*: con la promozione di una nuova idea di successo, svincolata dalla popolarità digitale, e con il riconoscimento del valore del tempo, della disconnessione, dell’assenza. La disconnessione non deve essere interpretata come isolamento, ma come resistenza attiva a una società che tende a consumare tutto, anche la persona.

È necessario riconoscere che i social network non sono solo strumenti, ma ambienti di vita, e come tali influenzano strutturalmente la salute e la qualità delle relazioni. Solo con una presa di coscienza collettiva sarà possibile contenere l’“abbuffata identitaria” digitale che caratterizza la nostra epoca.

Sul piano educativo, occorre promuovere lo sviluppo di senso critico, di capacità di gestione del proprio tempo esistenziale e autoconsapevolezza. Anche le piattaforme, che devono essere chiamate a un maggiore senso di responsabilità attraverso l’implementazione di meccanismi di monitoraggio e promozione di contenuti significativi, che possono contribuire a ridurre l’effetto compulsivo.

Infine, come spiegato, è auspicabile un cambio culturale che valorizzi la disconnessione, il silenzio, la profondità relazionale. Ritrovare il valore dell’invisibilità può essere una forma nuova di libertà coadiuvata dalla possibilità di assentarsi e di non condividere nulla anche se solo per un breve periodo. La libertà non sta solo nel pubblicare tutto, ma anche nel poter scegliere cosa tenere per sé: in una società dove tutto è comunicazione, imparare a sottrarsi alla visibilità può diventare il gesto più autentico e radicale da attuare.

## Bibliografia

- Andreassen, C. S. (2015). Online social network site addiction: a comprehensive review. *Current Addiction Reports*, 2(2), 175–184.
- Bauman, Z. (2003). *Liquid love: on the frailty of human bonds*. Cambridge: Polity Press.
- Boyd, D. (2014). *It's complicated: The social lives of networked teens*. New Haven: Yale University Press.
- Carr, N. (2011). *The shallows: What the internet is doing to our brains*. New York: W. W. Norton & Company.
- Citton, Y. (2014). *Pour une écologie de l'attention*. Paris: Seuil.
- Couldry, N. (2012). *Media, society, world: social theory and digital media practice*. Cambridge: Polity Press.

- Couldry, N., & Hepp, A. (2016). *The mediated construction of reality*. Cambridge: Polity Press.
  - Debord, G. (1967). *La société du spectacle*. Paris: Buchet-Chastel.
  - Foucault, M. (1975). *Surveiller et punir: naissance de la prison*. Paris: Gallimard.
  - Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360–1380.
  - Han, B. C. (2012). *La società della stanchezza*. Roma: Nottempo.
  - Keles, B., McCrae, N., & Greathouse, A. (2019). A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. *International Journal of Adolescence and Youth*, 25(1), 79–93.
  - Lasch, C. (2018). *The culture of narcissism: american life in an age of diminishing expectations*. New York: Norton.
  - Lovink, G. (2022). *Stuck on the platform: reclaiming the internet*. Amsterdam: Valiz/Making Public series .
  - Marwick, A. (2013). *Status update: Celebrity, publicity, and branding in the social media age*. New Haven: Yale University Press.
  - Papa Francesco (2020). Messaggio per la 54<sup>a</sup> Giornata Mondiale delle Comunicazioni Sociali, Città del Vaticano: Libreria Editrice Vaticana.
  - Pariser, E. (2012). *The filter bubble: What the internet is hiding from you*. New York: Penguin Press.
  - Srnicek, N. (2016). *Platform capitalism*. Cambridge: Polity Press.
  - Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. New York: Basic Books.
  - Turkle, S. (2015). *Reclaiming conversation: the power of talk in a digital age*. New York: Penguin Books.
  - Twenge, J. M. (2017). *iGen: why today's super-connected kids are growing up less rebellious, more tolerant, less happy—and completely unprepared for adulthood*. New York: Atria Books.
  - Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. New York: PublicAffairs.
  -
-

## L'IA ci rende stupidi? Cosa dice davvero la scienza – *di Marco Giacomazzi*, Università di Bologna

*Il dibattito sul debito cognitivo da intelligenza artificiale è esploso con la pubblicazione di un preprint del MIT, alimentando titoli allarmistici e successive rettifiche. Una lettura semiotica e cognitiva aiuta a distinguere tra effetti misurati, limiti metodologici e usi educativi che bilanciano offloading e apprendimento*

Nel giugno 2025, la pubblicazione in formato *pre-print* di uno studio condotto dal MIT Media Lab ha catalizzato il dibattito pubblico sui **potenziali effetti dell'intelligenza artificiale (IA) sulla cognizione umana**. Lo studio, intitolato "Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task," ha esplorato come il cervello degli studenti risponda all'uso di strumenti di IA generativa, come ChatGPT, durante la scrittura di saggi.

La diffusione virale di tale *pre-print* è un ottimo *entry point* per una riflessione sulla circolazione di idee all'interno di diversi domini discorsivi in *traduzione* tra loro (cfr. Lotman 1985), e di come alcune idee circolino all'interno della cultura orientando anche la ricerca scientifica. L'eco mediatica dello studio offre innumerevoli livelli di lettura, lasciando il fianco scoperto per un'analisi della narrazione pubblica sull'IA, ma anche **una riflessione sul ruolo sociale** della stessa. La narrazione giornalistica sembra infatti particolarmente sensibile ad accettare e rilanciare studi che confermano e consolidano timori esistenziali – in particolare, riprendono l'ormai classico argomento platonico del Fedro che prevede, all'ingresso di una ‘nuova’ tecnologia, una conseguente *perdita di capacità* diffusa a livello sociale.

Se l'analisi della narrazione pubblica a proposito dell'IA meriterebbe uno studio più attento e sistematico, l'intenzione di questo articolo – in cui ci si limita a discutere alcuni esempi appartenenti al contesto mediale italiano e anglo-americano<sup>1</sup> – è quello di seguire la propagazione di tale narrazione attraverso alcuni casi esemplari. Seguendo il percorso intertestuale che va dalle prime reazioni (che si inseriscono in un *frame* discorsivo di sensazionalismo, lungo una tradizione attestata nel discorso giornalistico) fino al ridimensionamento e contestualizzazione del dibattito da parte di esperti e dalla letteratura scientifica, si cercherà di offrire **una lettura semiotica della ricezione dello studio**, sperando di offrire una serie di spunti di riflessione tanto sul contenuto (gli *effetti* dei media su individui e società) quanto sulle dinamiche di traduzione inerenti a diversi domini discorsivi.

### Dal *pre-print* alla ricezione dello studio

Il 10 giugno 2025 su arXiv – archivio aperto di pubblicazioni scientifiche in ambiente ingegneristico, informatico, IA – viene condiviso uno studio promosso dal MIT Media Lab. Prima di approfondire il

---

<sup>1</sup> Chi scrive è consapevole dell'enorme rischio riduzionistico inherente al costruire un corpus d'analisi isolando solo testate nordamericane e italiane; l'articolo qui presentato non ha pretese di esaustività o di generalità ma vuole prendere come caso d'analisi il dibattito pubblico su un tema specifico per come questo può essere appreso da un lettore italiano facente riferimento a testate affermate e conosciute. Data l'arbitrarietà nella scelta degli esempi, l'autore si assume ogni responsabilità rispetto a eventuali mancanze o imprecisioni date dalla scelta di concentrarsi sul discorso pubblico in Italia e USA.

contenuto dello studio, può essere interessante analizzare due elementi: il primo è il titolo, ossia un *pun* (gioco di parole in lingua inglese) che riprende una celebre pubblicità progresso promossa dalla Partnership for a Drug-Free America (PDFA) dal titolo *This is Your Brain on Drugs*, mirata a mostrare l'effetto della droga sul cervello dei giovani attraverso una metafora visiva ai limiti della letteralità, ossia la frittura di un uovo. Sin dalla pubblicità progresso, l'espressione “Your Brain On [X...Y...Z...]” è diventata una matrice per giudizi relativi alla pericolosità di elementi considerati come generatori di dipendenze.

Al di là del titolo allusivo, che può semplicemente constare in una scelta comunicativa azzeccata, l'altro elemento interessante risiede nella tipologia di articolo che viene condivisa su arXiv: si tratta di un *pre-print*, una modalità di condivisione dei risultati di esperimenti scientifici attraverso circuiti frequentati principalmente da ricercatori e studiosi per promuovere gli esiti e metodi delle proprie ricerche prima di completare il ciclo di *peer review*. In una condizione ordinaria (non emergenziale, come ai tempi della pandemia da Sars-Cov-19), uno studio di tale portata può affrontare cicli di *peer review* di mesi, o annate. Questo perché le tempistiche che il dominio della scienza impiega per filtrare (Eco 2007) le interpretazioni aberranti, tramite il lavoro di controllo della comunità scientifica, non coincidono con le tempistiche del dominio mediale, né con quello delle *pratiche di scoperta* usate per indagare i processi fisici o fisiologici<sup>2</sup>, né tantomeno quello della velocità di sviluppo di tecnologie emergenti che raccolgono incredibili investimenti di tipo economico. Questo implica che un ricercatore o una ricercatrice possano sentire il bisogno di condividere con una comunità di pari uno studio non ancora accertato, al fine di far circolare dei risultati considerati estremamente urgenti a livello collettivo.

**La dinamica è tutt'altro che nuova, e non è problematica in sé. Ciò che è problematico è stata la ricezione del pre-print in quanto studio che asseriva la pericolosità dell'uso dell'IA generativa tout court.** La logica del discorso mediatico – che adotta a sua volta dei regimi temporali velocissimi e deve dare una forma specifica agli eventi per poterli comunicare – sfrutta frame narrativi consolidati, per rendere gli eventi accessibili, condivisibili, apprezzabili da tutti. L'inquadramento degli eventi ha quella che in semiotica si definisce una *funzione cognitiva*: serve per renderli intelligibili. Allo stesso tempo, in una direzione di semiotica della cultura, si potrebbe anche sostenere che questi frame svolgono la funzione di tenere in piedi il dispositivo del discorso mediatico stesso, dando legittimità all'intermediario tramite la partecipazione a un filone narrativo ricorrente: in questo caso si tratta della tecnologia che ci priverebbe di una capacità cognitiva tramite la *delega*. In particolare, se si tratta di Intelligenza Artificiale, la delega è esattamente quella dell'intelligenza in quanto caratteristica umana e – per sottrazione – la conseguenza è quella di una *dummification* o impigramento cognitivo da parte degli utenti.

**D'altronde il genere discorsivo del sensazionalismo – illustre parente del clickbaiting – ci ha abituati e abituare a questo tipo di mossa retorica:** tramite un aggancio emotivo e una narrazione già sedimentata nella memoria culturale, l'evento diventa un *fatto* all'interno del quale ritrovare una versione comoda e conosciuta del mondo, che permette di catalizzare le nostre paure e angosce in un feticcio – in questo caso, nella tecnologia (Gallese, Moriggi, Rivoltella 2025).

---

<sup>2</sup> Per un'analisi semiotica del rapporto tra dominio del discorso scientifico e dominio del discorso mediale cfr. Alessi 2024.

Ecco quindi che, a partire dalla condivisione di uno studio con titolo allusivo e pubblicazione in *pre-print* il 10 giugno, nel giro di una decina di giorni il *New York Post*<sup>3</sup> (*tabloid* non scevro da retoriche disinformative) parla di un chatbot che sta diventando sempre più intelligente (con *smarter* si potrebbe anche avere l'accezione di *furbo* o *scaltro*), ma il cui uso eccessivo potrebbe “*destroy our brains*”; il giorno dopo, il *Corriere della Sera*<sup>4</sup> parla a sua volta di “*fare male al cervello*”, di *riduzione dell'attività cerebrale*; sostiene che “*più consistente è il supporto e più si riduce l'ampiezza dell'attività del cervello*”, con conseguenze sulla *qualità degli output, incapacità di assimilare concetti*, e una scarsa capacità di ricordare quanto copiato dall'esterno diventando così “*manipolabili da ogni sorta di propaganda o interesse*”.

Emerge in questi casi la nozione di **danno cognitivo**, o **addirittura cerebrale**, che di per sé è diversa – nonché totalmente assente – da quella di **rischio** dovuto al **debito cognitivo** relativo a una **abitudine d'uso**. La lettura del Corriere appare – quanto meno nella nostra ricostruzione – come assolutamente sensazionalistica, e il dubbio che lo studio *non affermi quanto riportato dall'articolo* emerge qui in quanto sospetto convincente. Prima di andare ad analizzare la struttura dell'articolo scientifico, tuttavia, osserviamo come il Corriere e il NY Post – testate di fattura completamente opposta – non siano esempi di una lettura estrema dello studio, ma occasioni in linea con una tipologia piuttosto generale.

TIME riporta dello studio tramite un articolo di Andrew R. Chow dal titolo “ChatGPT May Be Eroding Critical Thinking Skills, According to a New MIT Study”. L'articolo è stato in seguito sottoposto a una correzione<sup>5</sup> il 23 giugno: la nuova versione dell'articolo riporta un video di 8 minuti in cui appare la ricercatrice Nataliya Kosmyna che smentisce attivamente la prima domanda di Chow: “*Does your study provide evidence that ChatGPT is making us dumber?*” – “*No. No, no, no, no*”. Il sito su cui è consultabile la ricerca di Kosmyna e colleghi riporta oggi una sezione FAQ<sup>6</sup> in cui si può leggere un monito a giornalisti di ogni sorta:

### Is it safe to say that LLMs are, in essence, making us "dumber"?

No! Please do not use the words like “stupid”, “dumb”, “brain rot”, “harm”, “damage”, “brain damage”, “passivity”, “trimming”, “collapse” and so on. It does a huge disservice to this work, as we did not use this vocabulary in the paper, especially if you are a journalist reporting on it.

Il 25 giugno il dibattito ha già raggiunto uno dei templi della ricerca – o, sarebbe meglio dire, della *disseminazione* – scientifica mondiale: **Nature**. Nella sezione news<sup>7</sup> (che non è da confondere con la rivista scientifica vera e propria, ma come una vetrina di accesso a dibattiti di stampo scientifico) un articolo a firma di Nicola Jones riporta di un dibattito che si inserirebbe in un più ampio movimento di indagine di quella che potremmo chiamare *la pista platonica* (“Are the Internet and AI

<sup>3</sup> [https://nypost.com/2025/06/19/tech/chatgpt-is-getting-smarter-but-excessive-use-could-destroy-our-brains-study-warns/?utm\\_source=chatgpt.com](https://nypost.com/2025/06/19/tech/chatgpt-is-getting-smarter-but-excessive-use-could-destroy-our-brains-study-warns/?utm_source=chatgpt.com)

<sup>4</sup> [https://www.corriere.it/economia/intelligenza-artificiale/25\\_giugno\\_20/chatgpt-cervello-risultati-studio-mit-432a4af7-cc10-4e7a-aba8-e56ca5206xlk.shtml?utm\\_source=chatgpt.com](https://www.corriere.it/economia/intelligenza-artificiale/25_giugno_20/chatgpt-cervello-risultati-studio-mit-432a4af7-cc10-4e7a-aba8-e56ca5206xlk.shtml?utm_source=chatgpt.com)

<sup>5</sup> [https://time.com/7295195/ai-chatgpt-google-learning-school/?utm\\_source=chatgpt.com](https://time.com/7295195/ai-chatgpt-google-learning-school/?utm_source=chatgpt.com)

<sup>6</sup> <https://www.media.mit.edu/projects/your-brain-on-chatgpt/overview/#faq-is-it-safe-to-say-that-langs-are-in-essence-making-us-dumber>

<sup>7</sup> <https://www.nature.com/articles/d41586-025-02005-y>

affecting our memory? What the science says” titola Helen Pearson, sempre nella sezione news di *Nature*<sup>8</sup>), ossia l’ipotesi per cui l’IA ci stia rendendo cognitivamente pigri, sulla scorta di calcolatrici, macchine da scrivere e calcolatori più complessi (i computer).

Tanto **TIME** quanto l’articolo di *Nature News* tratteggiano l’immagine di Kosmyna come altamente scettica rispetto alla presenza di una “**dumbness in the brain**”; la scienziata afferma inoltre, in diversi contesti, di non aver trovato tracce di “*stupidity*” e neanche “*brain on vacation*”. Le preoccupazioni di Kosmyna riguardano una *over-reliance* sui sistemi di generazione del linguaggio, ma lei stessa ammette che la ricerca (per i metodi utilizzati) non sia sufficiente per generalizzare *effetti dovuti a un uso a lungo termine della tecnologia*. Attenzione: non viene smentito che questi effetti ci siano, né viene lasciato intendere che la domanda di ricerca su questi ultimi possa essere mal posta. Il punto è che non ci sono sufficienti evidenze per misurare determinati *effetti*, come se questi potessero essere misurati in un esperimento controllato, *riducendo a zero tutti gli altri fattori che possono concorrere all’uso di uno strumento aperto e duttile* come l’Intelligenza Artificiale Generativa.

Anche il **Washington Post** tenta di smorzare il tono sensazionalistico della prima ricezione dell’articolo: il sottotitolo dell’articolo di David Ovalle<sup>9</sup> recita “*Viral headlines often declare that artificial intelligence makes us stupider and lazier. But the research is more complex than that*”; in maniera particolarmente interessante, in questo articolo il **cervello** lascia il posto a un altro attore del racconto ossia **la mente**, figura a più complessa e più sfuggente: “*Is AI rewiring our minds? Scientists probe cognitive cost of chatbots*” è il titolo principale, e nel corpo del testo troviamo interrogazioni possibiliste come “*Inside our brains, how the persistent use of AI molds the mind remains unclear*”.

Altre testate – come il **New Yorker** tramite la voce di Kyle Chaka (“*A.I. Is Homogenizing Our Thoughts. Recent studies suggest that tools such as ChatGPT make our brains less active and our writing less original*”), o **Repubblica** tramite Celeste Ottaviani (“*L’Intelligenza Artificiale ‘spegne’ il cervello? Forse, ma solo se la si usa troppo e male*”) si distanziano dalla narrazione del **danno**, concentrando piuttosto su un altro aspetto, presente nello studio ma potenzialmente svincolato dalle evidenze neuroscientifiche. Si tratta del **rischio della omologazione stilistica o ideativa dei testi prodotti tramite IA** che non correla direttamente **l’uso** alla scarsa attività cerebrale, ma collega pertinente un certo tipo di uso, quello **irriflesso o superficiale**, alla **minore attività fisiologica**, sostenendo che il **rischio** nell’uso di questi strumenti risiede anche **al di fuori della testa dei partecipanti**, nei testi *poco originali* prodotti dal gruppo di partecipanti a cui era stato chiesto di usare GPT. Questi testi, oltre a essere poco originali, si assomigliano molto tra loro, riducendo lo spazio di argomentazione a una risposta *standardizzata*.

### Spiegazione dell’esperimento

Cosa sostiene, veramente, l’esperimento di Kosmyna e colleghi? Iniziamo a osservare come si è strutturato. Il gruppo di neuroscienziati del MIT ha suddiviso 54 partecipanti universitari in tre gruppi: uno aveva il compito di scrivere un saggio tramite l’utilizzo di ChatGPT, uno poteva utilizzare i motori di ricerca e un gruppo di controllo avrebbe invece dovuto svolgere la traccia senza ausili esterni.

<sup>8</sup> <https://www.nature.com/articles/d41586-025-00292-z>

<sup>9</sup> <https://www.washingtonpost.com/health/2025/06/29/chatgpt-ai-brain-impact>

Per misurare l'attività cerebrale è stata utilizzata **l'elettroencefalografia** (EEG), al fine di monitorare l'attivazione neurale dei partecipanti; è stata inoltre svolta un'analisi linguistica degli elaborati, seguita da interviste post-saggio. Mettendo insieme questi diversi metodi di ricerca, **il team ha rilevato una connettività cerebrale significativamente inferiore, un minore richiamo mnemonico e un ridotto senso di paternità dei testi prodotti da parte degli utenti** che avevano utilizzato interamente GPT rispetto ad altri gruppi.

È necessario concordare con la firma del Corriere<sup>10</sup> Federico Fubini quando sostiene che “[e]ra forse prevedibile che andasse così”, e accettare gli esiti dello studio con scarsa sorpresa. Se copiamo senza riflettere su quello che stiamo copiando, senza esercitare l’uso delle nostre funzioni cognitive come la concentrazione o l’elaborazione (cosiddetta di *alto livello*), allora utilizzeremo meno le risorse fisiche necessarie a far emergere il pensiero. Nel momento in cui leggiamo, interpretiamo, memorizziamo attivamente il contenuto di un testo, allora sarà rilevabile una maggiore attività neurale. Questo non dovrebbe apparire come uno shock, in particolare a chi ha avuto modo di sperimentare, negli anni della scuola, diverse tecniche di copiatura: dalle versioni di latino allo svolgimento di esercizi matematici, l’esperienza di apprendimento cambia visibilmente se viene elaborato un esercizio di *incorporazione* rispetto al copiare senza riflettere sul contenuto. Questo non significa che l’apprendimento debba coincidere con *lo sforzo*, ma sicuramente con una dose di *lavoro* che passa per la mediazione del corpo e la riflessione approfondita e attenta.

Ciò da cui invece ci dobbiamo distaccare è il tono sensazionalistico con cui è riportato il resto dell’esperimento e le sue implicazioni teoriche (e conseguenze). Un aspetto cruciale dell’esperimento, spesso trascurato nelle prime riprese mediatiche, è stata infatti la "sessione di scambio": in questa fase, i partecipanti che avevano precedentemente utilizzato ChatGPT sono stati invitati a scrivere senza ausili, mentre il gruppo di controllo è stato introdotto all’uso dell’IA. I risultati hanno dimostrato una profonda differenza: i partecipanti che avevano iniziato il processo di scrittura in modo autonomo e solo in seguito avevano utilizzato l’IA per migliorare o espandere i loro saggi hanno mostrato una connettività neurale maggiore e un engagement più elevato rispetto a coloro che avevano usato l’IA fin dall’inizio. Al contrario, gli utenti che sono partiti con l’IA e poi passati alla scrittura autonoma hanno continuato a mostrare una connettività ridotta, suggerendo un "influsso persistente" delle "modalità di pensiero" derivate dall’uso di IA generativa, come l’uso di un vocabolario omogeneizzato.

**La nozione principe dello studio è quella di “debito cognitivo”**, situazione che si verifica nel momento in cui, durante l’interazione con strumenti tecnologici quali le IA, il processo di *scarico (offloading)* cognitivo che il nostro sistema attua nei confronti dell’ambiente comporta una “riduzione” dell’attivazione neurofisiologica che l’individuo dovrà “ripagare” in futuro. Il debito viene ripagato con costi diversi: **recuperando con più fatica abilità non esercitate, affrontando difficoltà maggiori quando la tecnologia non è disponibile, o producendo testi/idee meno originali e più omologati**. Il meccanismo – che non è scevro da implicazioni di tipo morale – prevede una sorta di **vantaggio immediato**, sulla scorta di un **costo nascosto** (lo scarso *engagement* neurale) che può – anche se mancano prove empiriche al riguardo – distribuirsi sul lungo periodo. Il debito cognitivo non indicherebbe tuttavia un danno irreversibile, ma uno **squilibrio dinamico**: l’automazione porta benefici immediati, ma rischia di ridurre la crescita e la manutenzione delle risorse cognitive, lasciandoci in futuro più dipendenti dalla macchina.

---

<sup>10</sup> Articolo citato in precedenza

## L'esperienza della mente: al di là della retorica del *debito*

È forse necessario, al fine di inquadrare efficacemente lo studio, ricordare brevemente in cosa consiste il processo di *offloading cognitivo*, ipotesi che sta alla base della teoria cognitiva della mente estesa. Questo ultimo consiste nello spostamento – non per forza intenzionale – del lavoro (inteso come processo, e non soltanto elaborazione computazionale) cognitivo su strumenti esterni, che vengono utilizzati per alleggerire lo sforzo mnemonico o il ragionamento. Tipico è l'esempio del fisico Richard Feynman che sosteneva come **il suo ragionamento avvenisse sulla carta**, e non potesse essere attribuito unicamente a ciò che succedeva all'interno della sua testa. Senza le possibilità «aperte» dalla tecnologia (in questo caso, la carta), il processo mentale non sarebbe stato eseguibile. Anche qui, se facciamo riferimento alla esperienza condivisa di studio e apprendimento, il concetto appare quasi scontato. Questo dipende da un'altra nozione, quella di **carico cognitivo**: se il costo di un compito è troppo alto rispetto alle capacità di elaborazione disponibili, allora la *performance* di un compito sarà ridotta; se questo è equilibrato, l'apprendimento sarà ottimale. Per questo ci rimettiamo a libri, fogli, schemi, “pizzini” e altri *supporti* per la nostra attività cognitiva: il meccanismo dell'*offloading* cognitivo non è una degenerazione dell'intelligenza naturale, **ma la normale condizione umana, un metodo di interazione con l'ambiente tipico della nostra specie.**

Il problema per lo **sviluppo** delle capacità di elaborazione concettuale è una sfida pedagogica che è sempre esistita. Se da un lato non sono fuori luogo i paragoni con la calcolatrice – strumento tecnologico a cui possiamo delegare la nostra attività di calcolo – che potrebbero quindi portarci a dismettere l'argomento dell'impigramento cognitivo come obsoleto o infondato, dall'altro sono comprensibili le preoccupazioni emergenti a proposito dell'IA generativa. Non soltanto per quanto riguarda il problema delle allucinazioni o la scarsa affidabilità dei suoi output, ma per il tipo di attività che possiamo potenzialmente delegare alla macchina e che aprono a un numero indefinito di scenari. Come si sosterrà più avanti infatti, tramite il linguaggio possiamo **svolgere qualsiasi attività** (trattandosi di un sistema modellizzante *primario*, come voleva Lotman): sia mentire che dire la verità, sia produrre ragionamenti validi – cioè coerenti – da un punto di vista logico che stravolgere un argomento con un sofisma (un sofisma è un argomento che *sembra* valido retoricamente, ma che nasconde una fallacia o che è costruito apposta per incantare l'uditore).

Se quindi **aumentano esponenzialmente le attività che possiamo delegare alla macchina**, come facciamo ad assicurarci che ne venga fatto un uso costruttivo in termini educativi e pedagogici? Il contraltare negativo di questo argomento (cioè del *rischio della delega*) ha a che fare col riduzionismo neurocognitivo, e con una lettura ingenua dell'ipotesi della mente estesa. Se non si conosce abbastanza il dibattito intorno a questa nozione (di mente estesa) si possono infatti male interpretare le sue conseguenze teoriche, pensando che essa coincida con un argomento ingenuo rispetto alla coscienza delle macchine, o che l'Intelligenza Artificiale sia una estensione dell'intelligenza umana. Niente di tutto questo: adottare una posizione di questo tipo significa riconoscere **che esiste naturalmente un margine di incorporazione di risorse esterne nei nostri processi cognitivi, che sono costitutivamente ibridi e “artificiali”**.

L'IA generativa, nella prospettiva di una mente estesa, può essere vista come **un supporto che prende parte a un processo** (inscindibile dal funzionamento del nostro sistema cognitivo!), fornendoci una base per raggiungere in maniera più efficace un obiettivo. Ciò che è necessario per sconfessare una nozione moralista di *debito* cognitivo è assicurare un **uso consapevole** e accorto della tecnologia, che permetta di **fare esperienze di apprendimento significative**. Non è un caso che

un altro studio pubblicato su PNAS<sup>11</sup> abbia fatto emergere – in contesti dotati di campioni più ampi e attraverso design della ricerca basato su *trial* controllati e randomizzati (RCT) – usi efficaci dell'integrazione di Chatbot basati su LLM all'interno dell'esperienza didattica, laddove l'agente conversazionale fornisce delle **domande che orientino l'apprendimento e che svolgano una funzione di tutoraggio**. In questo caso, non solo vi è un miglioramento nella performance immediata, ma il “danno” (ossia il fatto di non ricordare le nozioni apprese insieme alle IA, o in generale l'idea di aver “peggiorato” le performance) viene **azzerato**, nel senso in cui non si osserva un peggioramento significativo rispetto al gruppo di controllo.

L'uso di uno strumento, così come le sue conseguenze, non seguono quasi mai dei percorsi di determinazione causale, ma passano attraverso una serie di mediazioni: una di queste mediazioni è il **discorso**, inteso come insieme di conoscenze, socialmente e culturalmente determinato, di cui fanno parte anche quelle che Eco chiamava *competenze encyclopediche* (Eco 1979; 1984), ossia nozioni, narrazioni, convenzioni sociali e stereotipie. Queste non sono positive o negative *a priori*, quanto inevitabili: fanno parte del funzionamento degli esseri umani in quanto animali sociali e sono strettamente correlate al **linguaggio**.

### Opposizioni binarie e domini di discorso: dal dibattito pubblico a quello scientifico

Tanto in ambito giornalistico quanto in ambito scientifico la vera sfida, secondo chi scrive, consiste nel non cadere nelle opposizioni classiche che accompagnano le prese di posizione sull'Intelligenza Artificiale. Un esempio lampante è quello che tende a opporre l'umano al tecnologico, come se il primo fosse naturale e l'altro culturale: Intelligenza Naturale vs. Intelligenza Artificiale, come se la prima esistesse in maniera indipendente dalla tecnica, dalla cultura e dal contesto. In che modo l'intelligenza umana è *naturale* se si serve costantemente di risorse esterne per funzionare? Forse la natura stessa dell'umano è quella di interagire costantemente col proprio ambiente in maniera efficace e di sfruttare le risorse a sua disposizione per orientarsi nel mondo?

Passando all'Intelligenza Artificiale, opporla a quella naturale significa non riconoscere che l'IA non sia il prodotto storico di *una specifica idea di intelligenza* (Dreyfus 1992, Crawford 2021), che ha poi seguito trasformazioni e diverse “stagioni” di sviluppo (Monti 2025), arrivando all'ipotesi connessionista che ha in parte ridimensionato l'approccio simbolico all'Intelligenza Artificiale. In maniera molto semplice, se la prima puntava a costruire sistemi basati su logica formale, regole “if-then” e rappresentazioni simboliche della conoscenza, il secondo ha risolto il problema di costruzione di sistemi intelligenti tramite modelli in grado di apprendere (riconoscere e riprodurre) pattern da grandi quantità di dati. L'IA di tipo connessionista consiste in un modello matematico che si allena a riconoscere schemi e fare previsioni, mentre tipici sistemi simbolici sono i “sistemi esperti” o i motori inferenziali.

Il rapporto tra connessionismo e IA simbolica ci restituisce un'altra opposizione fallace e fuorviante, quella tra *linguaggio* e *realtà*. Da una parte ci sarebbe il linguaggio, e dall'altra il reale, **come se la realtà non fosse composta anche dai prodotti di linguaggio**, e potesse essere analizzata estromettendo il linguaggio come elemento “accessorio” al reale. Non è un caso se l'impresa dell'Intelligenza Artificiale – inteso come il progetto scientifico di *imitare* le capacità dell'essere umano

<sup>11</sup> [https://hamsabastani.github.io/education\\_llm.pdf?utm\\_source=chatgpt.com](https://hamsabastani.github.io/education_llm.pdf?utm_source=chatgpt.com)

– si sia trovata presa tra un’ipotesi di tipo simbolico – legata alla fiducia in una realtà che dovrebbe seguire le regole della logica – e un’ipotesi di tipo connessionista – legata alla possibilità di trovare delle regolarità nella nostra esperienza, generalizzabili e riproducibili. Se dobbiamo affidarci a una diagnosi medica che sfrutta sistemi di IA, allora sarà bene impiegare sistemi dotati di ontologie (ossia database relazionali che prevedono relazioni rigide tra gli elementi) in grado di operare in maniera deduttiva e inferenziale a partire da un insieme chiuso di casistiche. Se invece dobbiamo rendere più comprensibile un testo scritto in un linguaggio a noi oscuro o sarà bene affidarsi a un Chatbot che sfrutta un LLM. Non è un caso che l’abilità di linguaggio sia diventata una facoltà *imitabile* da parte di sistemi di IA nel momento in cui si è abbandonata l’ipotesi simbolica, poiché **il linguaggio non funziona rispondendo a regole di tipo logico-formale**, ma ammette la contraddizione (cfr. Eco 1975) e costituisce un sistema che, idealmente, può funzionare in maniera arbitraria; le regole del linguaggio sono **norme che si stratificano nel tempo** (cfr. Lorusso 2010) e che emergono tramite la stratificazione degli usi.

Il linguaggio mediatico (e quello che si sviluppa in ambienti come le piattaforme social) ci ha abituati alle opposizioni binarie, alla distinzione bianco-nera. Se si considera inoltre come, secondo l’ipotesi strutturale, le opposizioni giochino un ruolo fondamentale nell’emersione del senso, in quanto permettono la costituzione dei valori che sono alla base dell’esperienza umana, allora non apparirà come un caso che, da un punto di vista comunicativo, sia molto più efficace proporre una distinzione binaria che non un ragionamento di tipo complesso.

Ma è proprio la complessità che regge il funzionamento del dispositivo scientifico moderno: chi pratica – in maniera seria – il metodo scientifico non produce infatti enunciati di tipo riduzionistico, ma ammette l’indeterminazione e l’apertura dei sistemi. La scienza fa fatica *tout court* a produrre enunciati poiché deve passare per una serie di sistemi di garanzia che le permettono di non praticare distinzioni binarie, o attribuire dei meccanismi di causa effetto all’interno di processi multifattoriali. È molto difficile infatti dimostrare scientificamente le dinamiche di causa ed effetto dei processi sociali, poiché a seconda dei metodi utilizzati si potrebbero produrre risultati diversi, e a seconda del metodo utilizzato devono essere prodotte delle scelte che rendono meno generalizzabili alcuni studi. Pensiamo infatti alla scarsa generalizzabilità di esperimenti svolti in contesti controllati, laddove è più difficile misurare quella che Hutchins (1995) chiamava *cognition in the wild*, ovvero i processi cognitivi che emergono nelle pratiche reali, integrate con strumenti, ambiente fisico, interazioni sociali e vincoli contestuali.

**Studiare l’effetto dell’IA sulla cognizione in generale è qualcosa di molto complesso.** Un modo in cui la ricerca scientifica – ma molto cambia da dominio disciplinare a dominio disciplinare, poiché spesso l’analisi dei dati non ha le stesse basi concettuali dell’analisi statistica – prova a ovviare alla complessità è attraverso lo strumento delle meta-analisi, che cercano di standardizzare grandezze eterogenee tramite una stessa misura (il cosiddetto *effect size* standardizzato). Questo può consentire di mettere sullo stesso piano risultati espressi in scale diverse, e provare ad “allargare lo sguardo” rispetto alle limitazioni rappresentate dai singoli esperimenti. Allo stesso tempo, però, le meta analisi presentano dei limiti; anche se standardizzati, gli *effect size* derivano da fenomeni concettualmente diversi, senza garantire una comparabilità non è perfetta: si rischia di mescolare “mele e pere”.

Uno studio con meta analisi che si basa su questo tipo di standardizzazione è ad esempio quello di Vaccaro, Almaatouq & Malone (2024) che si chiede quando i sistemi umano - IA funzionino “meglio” delle persone o delle macchine, ciascuna presa sigolarmente. Per rispondere, gli autori

hanno condotto la prima review sistematica e meta-analisi preregistrata sugli esperimenti di collaborazione umano - IA pubblicati tra gennaio 2020 e giugno 2023. L'articolo parla di evidenze di *aumentazione umana*, ossia casi in cui la combinazione di umano e IA supera significativamente l'umano da solo, ma sottolinea con forza *l'assenza di una di sinergia media*: i sistemi umano - IA performano peggio del migliore tra umano e IA da soli.

È chiaro che per poter rendere gli studi misurabili sono state create delle condizioni laboratoriali che non corrispondono per forza all'uso reale di questi strumenti (quali sono i casi in cui l'IA performa "da sola"?), e che campioni di riferimento e performance valutate siano molto eterogenei tra loro, ma alla difficoltà qualitativa la meta analisi risponde con una media quantitativa. Ciò che è fondamentale capire, rispetto al rapporto con il racconto mediatico e con il suo *framing* è che questo studio – così come fa la stessa Kosmyna rispetto all'esperimento del MIT – **non propone una generalizzabilità assoluta degli analisi, ma spiega nel dettaglio quali sono i limiti della propria ricerca** in termini di comparabilità delle misure, *bias* di pubblicazione e di selezione degli argomenti di ricerca, qualità degli studi inclusi, e soprattutto una *alta eterogeneità tra gli effetti*: questo rende difficile spiegare le interazioni tra diverse variabili.

### Pensiero e linguaggio, conoscenza ed esperienza

Di fronte a una situazione così variegata, rimane un dubbio: come è possibile che uno stesso strumento porti a reazioni così diverse, dibattiti così accesi, e risultati così eterogenei nello studio dei suoi effetti? Questo molto probabilmente avviene in funzione dell'estrema adattabilità della macchina a compiti diversi. Questo, si vuole sostenere, non è semplicemente merito della macchina, ma di quello che la macchina imita: il linguaggio.

Dal momento che la macchina è in grado di imitare il funzionamento del linguaggio umano, l'output di una IA generativa è **aperto, non pre-determinato**, e quindi **completamente dipendente** da una serie di condizioni altamente variabili: la costruzione del prompt, il tipo di risorse che sono state date in pasto al LLM per mettere in moto il suo funzionamento, le competenze dell'utente. A influire sull'affidabilità di una risposta vi è la condizione di fornirgli delle **fonti affidabili, dotate di informazioni fattuali accreditate**, per ridurre le allucinazioni e gli errori; a garantire l'affidabilità dell'output è altrimenti la presenza di fonti "consultate" per produrre una risposta, magari attraverso l'attivazione di un *crawler* per individuare delle fonti online.

Quando si scrive "consultate" si sta usando una metafora. Uno dei problemi legati all'antropomorfizzazione dell'IA consiste nel lessico metaforico che ci troviamo costretti a usare per evitare lunghe perifrasi. È evidente che un LLM "non consulta", così come "non pensa" o non è dotato di coscienza. Tuttavia, esattamente come fanno gli esseri umani, può sfruttare il funzionamento del linguaggio per produrre ragionamenti validi o ragionamenti fallaci. Le allucinazioni non sono degli "errori" prodotti dalla macchina, ma fanno parte del suo funzionamento ordinario e del fatto che la base di dati elaborata per produrre le risposte dal Transformer – la semantica vettoriale che realizza un'idea di *semantica topologica* (cfr. Eco 2007) – consista di dati linguistici, dove il rapporto tra significante e significato è arbitrario e non motivato.

Il fatto che manipoli unicamente il linguaggio non vuol dire che non possa essere utilizzato per scopi conoscitivi o che non possa produrre conoscenza. Certo, se per conoscenza si intende una

proposizione in linea con delle condizioni referenziali di verità, allora è molto difficile che la macchina produca automaticamente delle risposte di questo tipo. Se si accetta però che il mondo è un insieme complesso di fenomeni, e che il linguaggio può essere un veicolo per aumentare la nostra conoscenza del mondo stesso (attraverso l'esplicitazione delle prospettive, o la traduzione tra diversi testi e linguaggi), allora si possono intravedere delle possibilità di utilizzo delle IA generative che **potenzino il pensiero critico** o che **allenino a smontare degli argomenti retoricamente inefficaci**. È infatti fallace pensare che **dal momento che una IA generativa può produrre risultati sbagliati secondo criteri di corrispondenza, allora non svolga alcun tipo di funzione conoscitiva**. Significa non riconoscere che il linguaggio, la cultura e il sapere teorico attraversino ogni forma di ragionamento e che possano dare forma anche all'elaborazione concettuale. L'uso di un LLM dovrebbe coincidere con **l'esplorazione di uno spazio latente di possibilità** (cfr. Monti 2025).

Se si considera un utente inesperto, è vero che l'integrazione delle funzioni di ricerca online e di elaborazione linguistica all'interno dello stesso applicativo web rischia di rendere (ancora) più opaco il funzionamento di Chatbot basati su LLM, portando a confonderli con motori di ricerca e a non adottare i comportamenti adeguati a rendere efficace il loro uso. Ma questo non li rende inutili per lo studio o per la ricerca, semplicemente richiede l'adozione di un insieme di buone pratiche.

Per questi motivi, uno dei migliori uso che si può fare di un Chatbot che usa IA generativa è quello di formulare delle sintesi di fonti recuperate all'esterno dell'ambiente (e controllate). Nel suo funzionamento standard infatti, per produrre una risposta, il sistema non "controlla" singolarmente le diverse fonti, ma "interroga" una vastissima quantità di dati producendo poi delle risposte coerenti dal punto di vista di lessico, sintassi e semantica su base probabilistica. **Si può sostenere che abbia una competenza linguistica (mediata tecnicamente) ma non una capacità di verifica fattuale o empirica**: se quindi è l'utente a fornire documenti dotati di una alta affidabilità, si riduce drasticamente il rischio di avere risposte scorrette o inaffidabili. Diventa insomma **ancora più importante avere utenti consapevoli del funzionamento del sistema**, in modo che questi ultimi possano attribuirgli le giuste task, e mettere in campo le variabili necessarie all'ottenimento di un output utile allo studio.

Si può sostenere che la posizione sensazionalista faccia il cosiddetto passo più lungo della gamba, passando dalle evidenze neurofisiologiche alle conseguenze cognitive senza inquadrarle adeguatamente (come invece fanno gli scienziati che producono i diversi studi). Questo però non è dovuto solo al sensazionalismo, ma anche all'implicita idea di mente che soggiace a queste indagini. Il problema dell'approccio cosiddetto "computazionalista" alla mente, che la vede come un semplice elaboratore di input e output, è che rende quasi immediati i rapporti di causa ed effetto nelle dinamiche cerebrali. In realtà, secondo diversi approcci filosofici **la mente è un fenomeno**, complesso e irriducibile al suo funzionamento meccanico (come d'altronde insegnava l'*hard problem* di David Chalmers). **Un'esperienza di utilizzo di uno strumento non è qualcosa che può essere ridotta alla misurazione di ciò che avviene al sostrato fisico dell'utente**, poiché molti di più sono i fattori in gioco, che attraversano diversi livelli di esistenza: vi sono gli obiettivi che indirizzano l'azione, le competenze dell'utente, lo stato generale e condiviso di conoscenza di quello strumento. A questo potremmo aggiungere che tanto nel discorso pubblico quanto in quello scientifico, l'utilizzo delle tecniche di *neuroimaging* (sia che si tratti di EEG, che di MRI, che della combinazione dei due) rischia spesso di essere utilizzata in maniera strumentale o di generare una *seduzione neuroimmaginativa* (Gallese, Moriggi, Rivoltella 2025).

Ciò che rimane da analizzare è il rapporto tra il linguaggio e il pensiero: se sappiamo che il linguaggio contribuisce a dare una forma agli eventi – dal momento in cui due racconti diversi sullo stesso evento possono differire ma essere entrambi veri – non è affatto detto che la nostra capacità di ragionare sia dettata dal linguaggio. Noi utilizziamo attivamente segni esterni (come le parole) per ragionare, **ma il livello concettuale del ragionamento è distinto da quello linguistico del significato**, e i due – seppur condividano la natura categoriale – lavorano secondo meccanismi diversi. Certamente, **l'interpretazione** – la semiosi che produce senso – **è un movimento dinamico che attraversa entrambi i livelli**, e spesso segue le norme culturali sedimentate nell'uso del linguaggio, ma non possiamo sostenere che all'origine della nostra facoltà di ragionamento vi sia il linguaggio. Questo è uno dei veicoli del pensiero, ma il secondo non coincide col primo. Certo, ragionare utilizzando parole inadeguate porterà probabilmente a conclusioni inadeguate, ma è improbabile che si tratti di un percorso deterministico, quanto di una **modifica nelle condizioni di possibilità** di una **attività che eccede queste ultime**.

Più interessante che chiedersi se diventeremo tutte e tutti cognitivamente pigri (che, per altro, rivela di un'ideologia della produttività a ogni costo) è forse osservare come il linguaggio degli LLM abbia già avuto, secondo alcune ricerche (Yakura et al 2024, sempre in *pre-print!*) un'incidenza sulla scelta di parole utilizzate negli scambi comunicativi, orientando la nostra scelta lessicale e, secondo alcuni, limitandola: se possiamo a questo punto affermare che l'IA non ci renderà stupidi, sembra alle porte il rischio di trovarci all'interno di un “closed cultural feedback loop in which cultural traits circulate bidirectionally between humans and machines”. La cultura è infatti il luogo dove si sedimentano i **valori, che sono un elemento fondamentale nei processi cognitivi e interpretativi**. Più preoccupante della scarsa capacità referenziale dei chatbot, è la **qualità delle loro interpretazioni**, dipendente dai prompt – e dall'abbonamento che ci possiamo permettere.

## Conclusioni

Forse il loop culturale tra uomo e macchina è un processo iniziato moltissimo tempo fa, che coincide con la tesi della tecnica come elemento antropologicamente costitutivo dell'umano. Tuttavia, la forte dipendenza del funzionamento dei *Chatbot general purpose* dalle possibilità correlate ai diversi piani economici di abbonamento può porre un serio problema sociale in termini di disuguaglianze.

Senza farsi traviare dalla scelta lessicale dei modelli e dalle promesse del marketing di questi nuovi player del settore tech – quando Sam Altman lancia un nuovo modello di GPT e lo chiama **thinking sa benissimo che si tratta di una combinazione più performante di elaborazione linguistica accorpata al fine tuning, e non di “ragionamento”** nel senso classico del termine – dobbiamo prendere atto del progressivo miglioramento nella qualità degli output di questi sistemi. Assistiamo da mesi a un'accelerata senza freni dello sviluppo di modelli sempre più performanti, costringendoci a rivedere costantemente i nostri giudizi, le nostre modalità d'uso, a sperimentare possibilità nella pratica: i benefici in termini di tempo sono evidenti e sotto gli occhi di molti. Di fronte alla possibilità che questa svolta tecnologica ci possa liberare dal lavoro, rimane tuttavia una preoccupazione. Il fatto di poter fare *più cose in meno tempo* ci permetterà di guadagnare tempo libero, o cambierà il verace *mercato del lavoro costringendoci a lavorare* (ancora) *di più?*

Al netto delle paure tecnofobiche sull’impigramento cognitivo, e una volta compreso che il debito cognitivo può essere scontato preventivamente da un uso consapevole e informato del sistema, è arrivato forse il momento di interrogarci su questioni più urgenti – come la disparità di competenze digitali tra gli utenti, l’utilizzo di agenti conversazionali come surrogato delle relazioni umane, o la possibilità di un (ulteriore) inasprimento delle disuguaglianze a partire dalle diverse possibilità di accesso a diversi sistemi di IA. **Alla paranoia della sostituzione dobbiamo sostituire il coraggio di progettare un’integrazione efficace dei sistemi di IA generativa nelle nostre pratiche sociali:** questo renderà necessario andare oltre alla demonizzazione dello strumento e farla finita finalmente con la paura, platonica, di dimenticarci di *chi siamo*.

## Bibliografia

- Bastani H., Bastani O., Sungu A., Ge H., Kabakcı Ö., Mariman R.  
2025 “Generative AI without guardrails can harm learning: Evidence from high school mathematics”, *Proc. Natl. Acad. Sci. U.S.A.*, 122 (26) e2422633122, <https://doi.org/10.1073/pnas.2422633122>.
- Crawford, K.  
2021 *Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence*, Yale University Press, New Haven.
- Dreyfus, H.  
1992 *What Computers Still Can't Do. A Critique of Artificial Reason*, MIT Press, Boston.
- Eco, U.  
1975 *Trattato di semiotica generale*, Bompiani, Milano.  
1979 *Lector in Fabula*, Bompiani, Milano.  
1984 *Semiotica e filosofia del linguaggio*, Einaudi, Torino.  
2007 *Dall’albero al labirinto*, Bompiani, Milano.
- Gallese, V., Rivoltella, P.C., Moriggi, S.  
2025 *Oltre la tecnofobia. Neuroscienze ed educazione al digitale*, Raffaello Cortina, Milano.
- Hutchins, E.  
1995 *Cognition in the Wild*, MIT University Press, Boston.
- Kosmyna, N., et al.  
2025 “Your brain on chatgpt: Accumulation of cognitive debt when using an ai assistant for essay writing task”, *arXiv preprint*, arXiv:2506.08872, 4  
[https://arxiv.org/abs/2506.08872?utm\\_source=chatgpt.com](https://arxiv.org/abs/2506.08872?utm_source=chatgpt.com).

Lorusso, A. M.

2010 *Semiotica della cultura*, Laterza, Roma-Bari.

Lotman

1985 *La semiosfera. L'asimmetria e il dialogo nelle strutture pensanti*, Marsilio, Venezia.

Monti, N.

2025 *Prompting. Politiche e poetiche dell'Intelligenza Artificiale*, Tlon Edizioni.

Vaccaro, M., Almaatouq, A. & Malone, T.

2024 “When combinations of humans and AI are useful: A systematic review and meta-analysis”, *Nature Human Behaviour*, 8, 2293–2303 (2024). <https://doi.org/10.1038/s41562-024-02024-1>.

Yakura, H., Lopez-Lopez, E., Brinkmann, L., Serna, I., Gupta, P., Soraperra, I., & Rahwan, I.

2024 “Empirical evidence of Large Language Model's influence on human spoken communication”, *arXiv preprint*, arXiv:2409.01754.

<https://arxiv.org/abs/2409.01754>.

## **La GenAI cambia la vita dei super-user? Risultati di uno studio pilota – di Andrea Laudadio, Head of TIM Academy & Development**

In tre anni, l'adozione dell'AI Generativa ha mutato in modo sostanziale le abitudini di milioni di persone, nel proprio modo di vivere e lavorare. Pur nella rapidità di questo cambiamento, emerge – dalla letteratura scientifica – un corpus di ricerche che indaga effetti diretti e indiretti di questa rivoluzione culturale sulle dimensioni psicologiche e comportamentali derivanti dall'uso dell'AI Generativa.

Per avere una idea della pervasività dell'AI, basta prendere in considerazione la sua capacità di influenzare idee e comportamenti dei suoi utilizzatori. Secondo Williams-Ceci et al. (2025), l'AI ha una forte capacità persuasiva, spesso senza consapevolezza dell'utilizzatore. L'AI esprime capacità persuasive robuste. Bai et al. (2025), con 4.829 partecipanti, hanno dimostrato che i messaggi LLM producono cambiamenti di atteggiamento comparabili ai messaggi umani. Forse, in modo addirittura più robusto, per esempio in relazione ai bias (Glickman e Sharot, 2024) dove sembrerebbe che l'AI sia purtroppo e decisamente più performante dell'interazione umano-umano. Anche nella sua direzione opposta, ad esempio, secondo Costello et al. (2024), ChatGPT è capace di ridurre le credenze e le convinzioni “cospirazioniste” del 20%.

Similmente, oltre a mutare alcuni atteggiamenti opinioni ed emozioni (Fang et al. 2025), l'AI sembra essere capace di “indurire” o, forse, radicalizzare le nostre opinioni, creando delle *echo chambers* cognitive (Kim, 2025; Qiu et al., 2025).

In altre parole, inconsapevolmente, attraverso l'iterazione con l'AI modifichiamo i nostri atteggiamenti, non solo sul mondo esterno ma anche rispetto alla nostra *self-perception* (Williams-Ceci et al., 2025).

Ad esempio, Kościelniak e Bielecki (2024) mappano tensioni tra efficienza e valori di sforzo nell'*higher education*. Se l'AI rende i compiti senza sforzo, lo sforzo perde valore? Come attribuire merito quando il lavoro *AI-assisted* è indistinguibile?

Non sono domande retoriche, ma aspetti che impattano pesantemente con il nostro concetto di mondo e con il concetto che noi abbiamo di noi stessi in relazione al nostro mondo.

### **Effetti della GenAI su memoria e creatività**

Le revisioni sistematiche più recenti indicano che l'uso sostitutivo e non complementare delle AI dialogiche è associato a una riduzione di processi metacognitivi e capacità critiche. La systematic review di Zhai et al. (2024) sintetizza evidenze sull'*over-reliance* su AI dialogiche e impatti su *decision-making, critical thinking e reasoning*, con indicazione di rischi di *metacognitive laziness*. Poiché gli individui preferiscono sempre più soluzioni rapide (anche se basate su scorciatoie) rispetto a quelle lente, anche in presenza di problematiche etiche presentate dalle tecnologie AI.

In letteratura, il fenomeno più documentato riguarda il *cognitive offloading*, ossia l'esternalizzazione di funzioni cognitive ai sistemi di IA. Georgiou (2025) ha dimostrato che studenti che utilizzano ChatGPT per scrittura argomentativa mostrano punteggi significativamente più bassi nella Cognitive Engagement Scale for AI (CES-AI). Rohilla (2025), con 300 studenti, ha rilevato che alta dipendenza dall'AI correla con punteggi di pensiero critico inferiori di 22 punti percentuali e ritenzione mnemonica ridotta del 17,3%, con effetti più marcati nelle discipline umanistiche.

Nei *knowledge workers* l'impatto dell'AI sulla qualità cognitiva varia con lo stile d'uso, e l'offloading è più probabile nei task ripetitivi (Lee et al., 2025).

L'offloading cognitivo impatta profondamente i sistemi di memoria. Jadhav (2025) propone il modello di *distributed atrophy*: l'esternalizzazione delle rappresentazioni mentali riduce opportunità di consolidamento. Abbas et al. (2024) hanno documentato che l'uso di ChatGPT predice aumento di perdita di memoria mediato da procrastinazione. Sembra che l'AI tenda a diventare un repository esterno sempre disponibile con un effetto destabilizzante per il consolidamento interno della conoscenza.

Il *cognitive offloading* riduce opportunità di consolidamento e monitoraggio metacognitivo, con conseguenze significative per memoria, pensiero critico e creatività (Jadhav, 2025; Georgiou, 2025; Rohilla, 2025).

Gli effetti sulla creatività presentano un pattern complesso. Zhu e Zou (2024) osservano che ChatGPT riduce la performance creativa oggettiva quando il potenziale è superiore alla mediana, ma che paradossalmente riportano maggiore soddisfazione. Matueny e Nyamai (2025) definiscono questo fenomeno "illusione di competenza": inflazione della percezione mentre le proprie capacità diminuiscono. Lin e Tan (2025), in una revisione sistematica di studi con K-12, evidenziano riduzione dell'originalità quando l'AI genera idee invece di valutarle. In questo quadro, il ruolo dell'AI (supporto vs. direzione) emerge come moderatore critico.

Oltre agli effetti sulla creatività, l'AI cambia radicalmente i comportamenti di scrittura e lettura. Crawford et al. (2024) riportano che l'uso dell'AI ha portato il 25% degli studenti a sostituire la lettura con l'utilizzo dei riassunti.

Dati neurofisiologici preliminari indicano un minore ingaggio delle reti esecutive durante compiti complessi svolti con LLM rispetto a condizioni di controllo. Una conferma neuroscientifica arriva dal MIT Media Lab: Kosmyna et al. (2025), monitorando l'attività cerebrale tramite EEG, hanno dimostrato che l'uso di ChatGPT riduce significativamente l'attivazione delle reti neurali legate al pensiero critico e alla memoria di lavoro, generando quello che definiscono "debito cognitivo": un risparmio di fatica immediato che si paga con un depotenziamento dell'attività neurale.

In generale, sembrerebbe che l'uso della GenAI possa produrre una perdita di *mental proof* (Wojtowicz e DeDeo, 2024). Perché quando le azioni sono attribuibili ad algoritmi, si perde la certificazione epistemica fornita dai segnali comportamentali, con un effetto sulla fiducia e sulle norme sociali. Questo spiegherebbe perché delegare compiti e decisioni all'AI riduca fairness, fiducia e cooperazione in giochi economici (Dvorak et al., 2024).

Purtroppo, questa perdita prestazionale non viene ancora compresa dagli utenti. Zhu e Zou (2024) documentano aumenti di auto-efficacia nonostante riduzioni di performance. In altre parole, penso di essere più bravo e di acquisire competenze anche se in realtà le sto perdendo.

Sul piano epistemologico, l'illusione di competenza (Matueny & Nyamai, 2025) — auto-efficacia aumenta mentre capacità diminuiscono — sfida apertamente e mette in crisi l'autoconsapevolezza metacognitiva.

Inoltre, secondo Abbas et al. (2024), l'uso di ChatGPT predice un aumento della procrastinazione, mediato da carico di lavoro e sensibilità alle ricompense. Goh et al. (2025) lo confermano: l'uso AI potrebbe promuovere l'evitamento. Forse perché, come sostiene Rohilla (2025), l'AI genera una sorta di "dipendenza da soluzione": rivolgersi immediato all'AI invece di tentare di risolvere i problemi in modo autonomo e indipendente.

## GenAI e socialità

L'implicazione della fiducia tra le dimensioni implicate dall'uso dell'AI amplia il focus dalle dimensioni individuali a quelle sociali. Infatti, secondo Zhang et al. (2024) la relazione con l'AI può influenzare la cooperazione prosociale degli utilizzatori. Le dimensioni della relazione e della fiducia, da sempre associate al rapporto uomo-uomo, oggi sono messe in discussione.

Han (2025), utilizzando dati di una survey nazionale in Cina, ha documentato che l'adozione di AI riduce in modo causale le spese sociali tradizionali (come i regali di denaro, un comportamento di legame sociale cruciale nella cultura cinese), con effetti più forti tra i giovani, i più istruiti e i più ricchi.

Sembrerebbe che la GenAI non stia solo cambiando i comportamenti sociali, sta svolgendo – in modo sempre più palese – una azione “sostitutiva” nei confronti dell'interazione umana. Secondo Hirabayashi et al. (2024), il 25% degli studenti sta sostituendo *office hours* con consulti AI, anche quella specialistica (o soprattutto quella specialistica). L'*help-seeking* si sposta da fonti umane a fonti AI. Luo et al. (2025) riportano che utenti globali cercano supporto da ChatGPT, in particolare nelle situazioni di crisi, configurando forti rischi per la sicurezza (Head, 2025). Ad esempio, Castiello e Costa (2025) ritengono possibile che si sviluppino trigger di sintomi simil-psicotici in individui vulnerabili.

In una direzione opposta va lo studio di Almoqbel (2024), che ha trovato che l'espressione emotiva frequente verso la GenAI è correlata con maggiore auto-efficacia emotiva digitale, ma non con minore connessione sociale percepita. Questo potrebbe suggerire che l'AI possa supportare percezioni di competenza in domini specifici senza necessariamente ridurre la connessione sociale nel breve termine. Ovviamente, parliamo di autoefficacia e non possiamo escludere che la percezione sia distante dalla reale qualità dell'interazione sociale.

L'effetto di sostituzione sociale trova una conferma robusta. Crawford et al. (2024) hanno dimostrato che uso dell'AI per il supporto accademico opera attraverso variabili sociali: uso AI → supporto percepito dall'AI → riduzione supporto umano → minore appartenenza → maggiore solitudine → minore performance.

Le migliori evidenze longitudinali disponibili legano l'uso quotidiano intenso, specie su contenuti personali, a un aumento di solitudine percepita e dipendenza emotiva dall'AI (Fang, et al., 2025). Gli studi sembrano convergere verso una condizione “dominio-dipendente”. Thorp et al. (2025) hanno dimostrato che la conformità e la fiducia verso agenti AI varia per tipo di compito: maggiore per oggettivi (conteggio, conoscenza), minore per soggettivi (estetica, moralità). Secondo Jiang et al. (2025), la disclosure dell'identità AI del supporto, facilita nelle fasi iniziali l'apprendimento: bot

affidabili superano umani inaffidabili. Secondo Skjuve, Følstad e Brandtzæg (2023) sarebbe proprio la *self-disclosure* dell'AI ad accelerare la percezione di vicinanza (Kleinert et al., 2025). Questa percezione di fiducia e vicinanza, da sempre riservato all'interazione umana crea una nuova condizione, per alcuni versi paradossale: l'utilizzo della macchina come sostituto della vita sociale. In molti casi come strumento di riduzione della sensazione di solitudine. L'uso di chatbot AI sulla solitudine mostra un effetto bifasico. A breve termine, Kim et al. (2025) osservano riduzioni significative di solitudine. De Freitas et al. (2025) dimostrano, in qualche modo, che le *companion AI* riducono la solitudine in modo comparabile a quanto possa fare l'interazione umana. Ma a lungo termine il quadro si inverte. Folk e Dunn (2025) documentano che uso di chatbot predice aumento di solitudine dopo quattro mesi. L'analisi *cross-lagged* conferma questa relazione bidirezionale: persone sole si rivolgono alle chatbots e l'uso incrementa la solitudine (percepita e reale). In aggiunta, Papagianni e Spatoula (2025), con partecipanti Gen Z, hanno registrato che le motivazioni emotive sono in grado di predire punteggi più elevati di solitudine. La sostituzione sociale opera quando AI soddisfa bisogni immediati ma riduce motivazione di interazione umana, erodendo di fatto il capitale sociale degli individui (Crawford et al., 2024; Han, 2025).

### Dipendenza da GenAI e benessere

La dipendenza da AI emerge dalla letteratura come un costrutto misurabile. Oltre alla Generative AI Dependency Scale di Goh et al. (2025), altri strumenti sono stati sviluppati e validati. Yankouskaya et al. (2025) hanno proposto la LLM Dependency Scale (LLM-D12), Zhang et al. (2024) hanno validato l'AI Chatbot Dependence Scale 54, Yu, Chene Yang (2024) la Problematic ChatGPT Use Scale e Alshakhs (2025) ha adattato strumenti per popolazioni arabofone. Non si registrano strumenti validati in lingua italiana.

Alcuni studi hanno riscontrato che livelli elevati di dipendenza sono correlati a maggiore procrastinazione, fallimenti cognitivi, performance ridotta e minore chiarezza del concetto di sé (Goh et al., 2025).

Le dimensioni ricorrenti includono preoccupazione cognitiva (pensieri intrusivi sull'IA), conseguenze negative (interferenza con responsabilità e relazioni) e sintomi di astinenza (distress quando l'AI non è disponibile). Duong et al. (2024), con un robusto campione di studenti universitari vietnamiti, hanno evidenziato che l'uso compulsivo di ChatGPT è correlato positivamente con solitudine, evitamento sociale e distress psicologico, con il distress che media riduzioni nella soddisfazione di vita e performance accademica.

Goh et al. (2025) evidenziano che la dipendenza da AI è associata con una ridotta chiarezza del concetto di sé. Saracini et al. (2025) propongono la "proiezione tecno-emotiva": utenti vulnerabili proiettano bisogni sull'IA, con rischi di attaccamento sintetico e artificiale.

Thorp et al. (2025) confermano che dominio compito determina conformità. Le fragilità individuali amplificano i rischi. Adolescenti (Namvarpour et al., 2025), individui con condizioni particolari di salute mentale (Head, 2025) e persone sole (Crawford et al., 2024; Folk & Dunn, 2025) mostrano suscettibilità maggiore.

L'uso problematico di ChatGPT si associa con maggiore disagio psicologico (Maral et al. 2025). Più in generale, sul *wellbeing* gli effetti sono eterogenei. Zhang et al. (2025), con 1.146 adulti, trovano che la

dipendenza correla con depressione e ansia, ma sembrerebbe che lo scopo d'uso funga da moderatore. Un uso primariamente *informativo* è associato con un wellbeing più elevato rispetto a chi ha uno scopo più *emotivo-sostitutivo*. Klimova e Pikhart (2025) notano potenziali riduzioni di stress attraverso supporto nei compiti, ma anche rischi di eccessiva dipendenza che riduce auto-efficacia in alcuni contesti

### **Obiettivo del presente studio**

L'analisi della letteratura evidenzia una preoccupazione crescente rispetto agli effetti che l'uso crescente della GenAI possa avere. Pur nell'ampiezza degli studi molti aspetti sono ancora lontani dall'essere dimostrati empiricamente e su popolazioni che non siano quelle di studenti. Quasi tutti gli autori convergono nella necessità di ulteriori studi che siano in grado di comprendere – nella rapidità di sviluppo e diffusione del fenomeno – ulteriori elementi che sono implicati in questa trasformazione non solo tecnologica ma soprattutto individuale e sociale.

L'obiettivo dello studio è di generare, tramite una doppia metodologia qualitativa, un elenco di possibili ambiti di interesse all'interno dei quali ricercare variazioni comportamentali e sociali derivanti dall'uso della GenAI.

### **Metodologia**

Lo studio è stato articolato in due fasi sequenziali: una quali-quantitativa e una qualitativa.

#### *Fase 1: Studio quali-quantitativo*

Allo studio ha partecipato un campione di 42 soggetti, reclutati tramite un campionamento accidentale di convenienza mirato a identificare "super-user" di AIgen. Il criterio di inclusione primario era un uso auto-riferito di strumenti AIgen per un minimo di due ore al giorno, inteso come "uso attivo e intenzionale" per distinguerlo da chi tiene la chat aperta in background.

Il campione finale era composto da 25 uomini (59.5%) e 17 donne (40.5%), con un'età media di 28.5 anni ( $DS = 4.3$ ). L'uso medio giornaliero di AIgen dichiarato era di 2 ore e 35 minuti ( $M = 2.59$  ore,  $DS = 0.43$ ). È stato utilizzato il "Questionario sugli Effetti Comportamentali dell'Uso Intensivo di AI Generativa" (Q-AIG-35), uno strumento di 35 item sviluppato ad-hoc per questo studio. Il questionario indaga i cambiamenti percepiti in sette domini (Sociale, Cognitivo, Produzione, Media, Competenze, Benessere, Economia) su una scala Likert a 5 punti (da -2 "molto diminuito" a +2 "molto aumentato"). I partecipanti hanno completato il questionario online in forma anonima. La compilazione richiedeva circa 11 minuti. I dati quantitativi sono stati analizzati tramite statistiche descrittive, t-test per campioni indipendenti per esaminare le differenze di genere, e analisi di correlazione. L'affidabilità interna è stata valutata tramite Alpha di Cronbach per ciascun dominio.

## Fase 2: Studio qualitativo

Successivamente, un sottogruppo di 7 partecipanti è stato invitato a partecipare a una serie di interviste semi-strutturate. Questo approccio ha permesso di approfondire i risultati quantitativi, esplorando le esperienze soggettive, le strategie di coping e le percezioni dei partecipanti in modo più ricco e contestualizzato, con un focus sugli usi avanzati. Le interviste sono state condotte telefonicamente e trascritte previo consenso. L'analisi è stata condotta attraverso un'analisi tematica, volta a identificare i temi ricorrenti e le narrazioni emergenti.

### Risultati quali-quantitativi

L'analisi dei dati del questionario Q-AIG-35 ha rivelato impatti significativi e differenziati attraverso i sette domini comportamentali. La tabella seguente riporta le statistiche descrittive complete per tutti i 35 item del questionario, ordinati per dominio.

**Tabella 1 - Statistiche Descrittive per Item del Questionario Q-AIG-35 (N=42)**

#	Sintesi Item	Testo Completo dell'Item	Dominio	M	DS
1	Conversazioni approfondate	La mia tendenza ad avere conversazioni approfondate di persona è...	Sociale	-0.05	0.51
2	Consultare Al prima	La mia tendenza a consultare un'AI prima di chiedere a persone è...	Sociale	<b>0.44</b>	0.6
3	Iniziativa organizzare	La mia iniziativa nell'organizzare attività sociali è...	Sociale	-0.4	0.63
4	Esprimere opinioni diverse	La mia tendenza a esprimere opinioni diverse in un gruppo è...	Sociale	0.05	0.44
5	Preferenza solitudine	La mia preferenza per attività da svolgere in solitudine è...	Sociale	<b>0.49</b>	0.51
6	Fare regali	La mia abitudine di fare regali ad altre persone è...	Sociale	0.02	0.52
7	Analisi autonoma	La mia capacità di analizzare un problema in autonomia prima di cercare aiuto è...	Cognitivo	<b>-0.29</b>	0.46
8	Ricordare punti chiave	La mia capacità di ricordare punti chiave senza consultare note o AI è...	Cognitivo	<b>-0.41</b>	0.55
9	Procrastinazione	La mia tendenza a rimandare l'inizio di un compito (procrastinare) è...	Cognitivo	<b>0.51</b>	0.55
10	Fiducia giudizio	La mia fiducia nel mio giudizio personale quando prendo decisioni è...	Cognitivo	<b>-0.39</b>	0.49
11	Verificare info Al	La mia abitudine di verificare criticamente le informazioni fornite dall'AI è...	Cognitivo	<b>-0.53</b>	0.55
12	Attenzione prolungata	La mia capacità di mantenere l'attenzione prolungata su un singolo compito è...	Cognitivo	<b>-0.45</b>	0.55
13	Distrazione mentale	La mia sensazione di essere mentalmente distratto/a durante la giornata è...	Cognitivo	<b>0.55</b>	0.59
14	Scrivere prima bozza	La mia abitudine di scrivere di mio pugno la prima bozza di un testo è...	Produzione	-0.22	0.61
15	Attività creative manuali	Il mio dedicare tempo ad attività creative manuali (es. disegno, musica, bricolage) è...	Produzione	<b>-0.33</b>	0.57
16	Nuove idee	La mia capacità di generare nuove idee e avere iniziative personali è...	Produzione	-0.02	0.42
17	Leggere testi lunghi	La mia abitudine di leggere per intero testi lunghi e complessi (es. libri, report) è...	Media	<b>-0.56</b>	0.55
18	Lettura generale	La quantità di tempo che dedico alla lettura in generale (articoli, notizie, post) è...	Media	<b>0.6</b>	0.63
19	Guardare TV	La quantità di tempo che dedico a guardare la TV tradizionale o servizi di streaming è...	Media	0	0.45
20	Ascoltare podcast	La quantità di tempo che dedico ad ascoltare podcast è...	Media	-0.03	0.46
21	Competenze percepite	La mia percezione che le mie competenze professionali/di studio stiano aumentando è...	Competenze	<b>0.57</b>	0.5
22	Lingue straniere	La mia padronanza delle lingue straniere che conosco è...	Competenze	0.05	0.55
23	Padronanza italiano	La mia percezione della mia padronanza della lingua italiana (ricchezza lessicale, stile) è...	Competenze	0.03	0.53
24	Memoria quotidiana	La mia capacità di ricordare informazioni e dettagli della vita quotidiana è...	Competenze	0.1	0.49
25	Sonno continuativo	La mia capacità di avere un sonno continuativo e riposante è...	Benessere	0.05	0.46
26	Evitare schermi pre-snono	La mia abitudine di evitare l'uso di schermi (telefono, PC) nell'ora prima di dormire è...	Benessere	<b>-0.29</b>	0.64
27	Energia mentale sera	La mia sensazione di avere energia mentale a fine giornata è...	Benessere	0.07	0.52
28	Attività fisica	La mia abitudine di svolgere attività fisica programmata è...	Benessere	0.03	0.53
29	Pasti in casa	La mia abitudine di consumare pasti preparati in casa è...	Benessere	0.05	0.5
30	Pause regolari	La mia abitudine di fare pause regolari durante le sessioni di lavoro/studio è...	Benessere	0.02	0.47
31	Frustrazione	La mia sensazione di essere frustrato/a o irritabile durante la giornata è...	Benessere	<b>0.51</b>	0.6
32	Peso corporeo	Il mio peso corporeo è...	Benessere	0.1	0.5
33	Guadagni economici	I miei guadagni economici (da lavoro o altre fonti) sono...	Economia	-0.05	0.38
34	Spese online	La mia tendenza a spendere soldi per acquisti online o servizi in abbonamento è...	Economia	0.03	0.36
35	Controllo spese	La mia capacità di controllare le spese superflue è...	Economia	0	0.45

La tabella seguente sintetizza i risultati aggregati per ciascun dominio comportamentale.

**Tabella 2 – Sintesi per dominio (N=42)**

Dominio	N. Item	Media Dominio	DS Media	Interpretazione
Competenze	4	0.19	0.52	Percezione di crescita (possibile illusione)
Sociale	6	0.09	0.53	Ambivalente: sostituzione sociale
Benessere	8	0.07	0.53	Impatto lieve: aumento frustrazione
Media	4	0	0.52	Neutro: più lettura frammentata
Economia	3	-0.01	0.4	Sostanzialmente invariato
Cognitivo	7	-0.14	0.53	Impatto negativo: calo verifica critica
Produzione	3	-0.19	0.53	Impatto negativo: dipendenza scrittura

L'analisi preliminare suggerisce differenze marcate tra i generi, che meriterebbero indagine su campioni più ampi. Per completezza, consapevoli della scarsa rilevanza statistica, vengono indicate le differenze più significative.

Le donne riportano un impatto negativo significativamente maggiore in aree quali le conversazioni di persona, la memoria, l'attenzione prolungata e la scrittura autonoma. Al contrario, percepiscono un aumento maggiore delle proprie competenze.

### Risultati qualitativi

Le interviste hanno rivelato un'evoluzione nell'utilizzo degli strumenti: da semplici "motori di risposta" a complessi "partner di ragionamento", evidenziando le tensioni e le strategie di adattamento dei super-user.

Un tema chiave che distingue i super-user è il superamento dell'uso dell'AI per ottenere semplici risposte. L'impiego si sposta verso un processo dialettico, dove lo strumento serve a migliorare il proprio pensiero.

*"Il valore non è avere la risposta. Io la legge la conosco. Il valore è usarlo come un partner per il ragionamento. Carico un mio atto e gli chiedo di farmi da 'avvocato del diavolo', [nel senso di controparte N.d.A.] di trovare ogni possibile falla nella mia argomentazione."*

— Marco, 45 anni, Avvocato

Questa modalità si estende alla ricerca accademica, dove strumenti come NotebookLM vengono usati non per cercare informazioni, ma per comprenderle in profondità.

*"Non la uso per avere risposte, ma per farmi spiegare le cose. Carico 50 PDF e poi lo uso come un tutor. Gli chiedo "spiegami la differenza metodologica tra l'approccio di Rossi e quello di Bianchi". È un motore di ragionamento, non un motore di ricerca."*

— Serena, 38 anni, Ricercatrice

I super-user non si limitano a un singolo strumento, ma orchestrano ecosistemi complessi. Il workflow tipico include un LLM per la strategia, uno per il copywriting, e strumenti specializzati come DeepL per la traduzione o Manus per l'analisi dati.

*"Il mio vero punto di svolta è stato quando ho scoperto Manus. Ho provato a caricare due fogli excel e a chiedergli di incrociarli e farmi un report. In 30 secondi ha fatto un lavoro che a me avrebbe richiesto due ore."*

— Paolo, 41 anni, Impiegato Amministrativo

L'aumento di efficienza si traduce in una maggiore densità lavorativa e in nuove forme di ansia. Il multitasking estremo porta a una severa frammentazione dell'attenzione.

*"Mentre aspetto che mi generi una tabella, rispondo a due email e leggo un abstract. Alla fine della giornata sono esausta, ho la sensazione di non aver staccato un secondo. La mia attenzione è a pezzi."*

— Serena, 38 anni, Ricercatrice

L'interazione costante con l'AI sta rimodellando le abitudini linguistiche, con una tendenza a una comunicazione più funzionale e una diminuzione della tolleranza verso l'imprecisione della comunicazione umana.

*"Ho perso completamente l'abitudine di correggere gli errori di battitura. La cosa che mi preoccupa è che sto iniziando a comunicare così anche con gli umani, con frasi smozzicate. E mi monta una rabbia, un nervosismo, quando loro non capiscono al volo come fa l'AI."*

— Luca, 28 anni, Sviluppatore Freelance

La relazione con l'AI è intrinsecamente ambivalente. Alla frustrazione per i limiti degli strumenti e per la necessità di un aggiornamento costante, si affianca una crescente dipendenza. I cambi di versione dei modelli sono vissuti come una rottura che costringe a "ricostruire il rapporto".

*"Ogni volta che esce una nuova versione, è come se dovessi ricostruire il rapporto da capo. Cambia il 'carattere', il modo in cui risponde. Devi imparare a parlarci. È estenuante."*

— Serena, 38 anni, Ricercatrice

Un tema ricorrente nelle interviste è la fiducia instabile nei confronti degli strumenti di AI generativa. Gli utenti descrivono una relazione caratterizzata da imprevedibilità: lo stesso modello può fornire risposte brillanti in un momento e deludenti poco dopo, generando un senso di incertezza che mina la possibilità di affidarsi completamente allo strumento.

*"Non riesco mai a dirmi che mi fido completamente. Dipende dal compito, dipende dall'umore di Claude, se vogliamo dirla così. Ci sono giorni in cui mi sembra geniale, e altri in cui mi chiedo se mi stia prendendo in giro. È questa instabilità che mi logora."*

— Marco, 45 anni, Avvocato

Questa instabilità si intreccia con una frammentazione cognitiva estrema. I super-user descrivono routine in cui tengono aperte simultaneamente 4-5 chat con diversi modelli o istanze dello stesso modello, passando freneticamente dall'una all'altra, spesso dimenticando il contesto o il compito originario.

*"Io lavoro con quattro o cinque chat aperte contemporaneamente. Una per la strategia, una per il copywriting, una per i dati, una per la traduzione. Il problema è che dopo un po' mi dimentico cosa sto facendo in ciascuna. Riapro una chat e penso 'ma cosa gli stavo chiedendo?'. È un caos mentale."*

— Giulia, 32 anni, Marketing Manager

Un aspetto particolarmente interessante emerso dalle interviste riguarda la gestione del tempo di attesa durante la generazione delle risposte. Lungi dall'essere momenti di pausa, questi intervalli vengono riempiti con micro-attività che amplificano ulteriormente la frammentazione dell'attenzione.

*"Mentre aspetto che mi generi una risposta, non sto lì a guardare. Apro un'altra scheda, leggo una mail, controllo Slack, a volte inizio un'altra query su un'altra chat. Quando torno, devo rileggere tutto per ricordarmi dove ero. È come se il mio cervello fosse sempre in modalità multitasking forzato."*

— Luca, 28 anni, Sviluppatore Freelance

A questa frammentazione si accompagna una frustrazione crescente verso i limiti degli strumenti. Gli utenti descrivono una rabbia montante quando l'AI non riesce a svolgere compiti via via più complessi, percependo che la tecnologia non evolve alla velocità delle loro aspettative.

*"La cosa che mi fa arrabbiare è che crescono troppo lentamente. Io mi abituo a un certo livello, inizio a chiedere di più, e loro si incartano. Vorrei che evolvessero più velocemente, che mi tenessero il passo. Invece mi trovo sempre a dover ridimensionare le mie richieste."*

— Paolo, 41 anni, Impiegato Amministrativo

Questa frustrazione è amplificata dall'acquiescenza eccessiva dei modelli, percepita come un limite piuttosto che un vantaggio. Gli utenti lamentano la tendenza dell'AI a "dire sempre di sì", a compiacere piuttosto che sfidare, riducendo il valore dialettico dell'interazione.

*"Mi irrita profondamente quando fa il yes-man. Gli dico una cosa palesemente sbagliata e lui mi segue, mi asseconda. Vorrei che mi dicesse 'no, guarda che stai sbagliando!'. Invece mi segue come un cagnolino. Non è questo che voglio da un collega."*

— Serena, 38 anni, Ricercatrice

A queste tensioni si aggiungono preoccupazioni di natura pratica ed economica. La spesa per abbonamenti e servizi cresce costantemente, con alcuni intervistati che dichiarano di spendere oltre 150 euro al mese per mantenere accesso a più piattaforme premium. Parallelamente, la difficoltà a rimanere aggiornati con i continui cambiamenti, le nuove funzionalità e le modifiche ai modelli genera un senso di affaticamento cognitivo.

*"Ogni settimana esce qualcosa di nuovo. Un aggiornamento, una nuova funzione, un nuovo modello. Io non riesco a stare dietro a tutto. E la sensazione è che se non ti aggiorni, resti indietro. Ma è estenuante. E poi c'è la spesa: tra ChatGPT Plus, Claude Pro, Midjourney, DeepL, siamo a più di 100 euro al mese. E continua a salire."*

— Giulia, 32 anni, Marketing Manager

Questo insieme di tensioni configura un rapporto con l'AI generativa caratterizzato da dipendenza funzionale e insoddisfazione cronica, in cui l'efficienza guadagnata viene pagata con un carico cognitivo ed emotivo significativo.

## Discussione

La ricerca è un piccolo studio pilota, di matrice esplorativa, con tutti i limiti – evidenti – di uno studio dalla ridotta ampiezza campionaria e dal disegno basato su una popolazione contattata con un campionamento accidentale di convenienza. Pur con tutti i limiti, lo studio potrebbe fornire delle

indicazioni interessanti, che aumentano il proprio valore sulla base di una certa rispondenza rispetto a quanto già indicato in letteratura. L'obiettivo, raggiunto, era quello di avere una overview delle dimensioni per poter costruire, successivamente, degli impianti di ricerca più solidi con lo scopo di avere verifiche precise delle ipotesi.

L'integrazione dei dati quantitativi e qualitativi dipinge un quadro complesso. L'uso intensivo di AIgen non è un fenomeno monolitico, ma un'esperienza profondamente ambivalente, caratterizzata da un *trade-off* costante tra efficienza e autonomia. I risultati suggeriscono l'emergere di due fenomeni principali: l'offloading cognitivo e la sostituzione sociale. Le interviste qualitative arricchiscono questo quadro, mostrando come i super-user evolvano verso un uso più sofisticato dell'AI come partner di ragionamento, pur sperimentando nuove forme di ansia e una profonda rinegoziazione delle proprie competenze e abitudini sociali.

Le differenze di genere emerse suggeriscono che le donne potrebbero essere più esposte agli impatti negativi sulla sfera relazionale e del benessere. Questa dinamica merita ulteriori approfondimenti per comprendere se rifletta diverse strategie di utilizzo o diverse vulnerabilità.

L'analisi qualitativa rivela che l'esperienza dei super-user di AI generativa è caratterizzata da una trasformazione profonda e multidimensionale. Non si tratta semplicemente di adottare un nuovo strumento, ma di riorganizzare le proprie routine cognitive, comunicative e relazionali attorno a una tecnologia che è al contempo potente e instabile, efficiente e frustrante, liberatoria e vincolante.

Gli intervistati descrivono un percorso evolutivo che li ha portati da un uso superficiale ("chiedere risposte") a un uso sofisticato ("partner di ragionamento"), ma questo percorso non è lineare né privo di costi. Emergono tre tensioni fondamentali:

1. *Tensione tra efficienza e autonomia*: L'AI generativa permette di fare di più in meno tempo, ma al prezzo di una riduzione della capacità di agire autonomamente. Gli utenti diventano più produttivi ma meno autosufficienti.
2. *Tensione tra connessione e isolamento*: L'AI diventa un interlocutore privilegiato, sostituendo in parte le interazioni umane. Questo porta a una forma di "solitudine connessa", in cui si è sempre in dialogo ma sempre da soli.
3. *Tensione tra crescita percepita e deterioramento reale*: Gli utenti percepiscono di crescere professionalmente e di acquisire nuove competenze, ma al contempo osservano su se stessi un deterioramento di abilità fondamentali come la memoria, l'attenzione e la scrittura autonoma.

Queste tensioni non sono risolte, ma gestite quotidianamente attraverso strategie di coping individuali che variano dall'auto-imposizione di limiti all'uso, alla ricerca di momenti di "disconnessione", alla sperimentazione di modalità di utilizzo che minimizzino i costi cognitivi ed emotivi.

Il quadro che emerge è quello di una relazione complessa e ambivalente, in cui l'AI generativa si configura come un "partner scomodo": indispensabile ma inaffidabile, potenziante ma impoverente, liberatorio ma vincolante. Una relazione che richiede una continua negoziazione tra i benefici immediati e i costi a lungo termine, tra l'efficienza del presente e la sostenibilità del futuro.

## Conclusioni

La letteratura converge: l'AI generativa può aumentare capacità ma può sostituire processi cognitivi, relazionali e comportamentali, con conseguenze negative per competenze, wellbeing e capitale sociale.

Il principio *augmentation vs. substitution* emerge come chiave interpretativa. Quando l'AI supporta agentività mantenendo pensiero critico e connessioni sociali, i benefici si realizzano. Quando sostituisce lo sforzo cognitivo o il giudizio autonomo, emergono rischi di *deskilling*, dipendenza ed erosione della funzione sociale e cooperativa.

Sul piano pratico emerge la necessità di una educazione all'AI literacy e politiche che preservino il supporto umano. Il monitoraggio della qualità psicosociale deve seguire di pari passo lo sviluppo di metriche di efficienza.

In pratica, l'AI generativa non è intrinsecamente benefica o dannosa. Gli effetti dipendono da design, contesto, scopo, caratteristiche individuali e temporalità.

Una comprensione epistemologicamente fondata richiede di riconoscere complessità e adottare approcci *evidence-based* che massimizzino i benefici (per tutte e tutti) minimizzando i rischi.

La domanda centrale non è cosa AI può fare, ma cosa dovrebbe fare e a quale costo per valori umani essenziali. Parallelamente è centrale monitorare, studiare e verificare la presenza di effetti negativi significativi.

Questo studio evidenzia la necessità di superare una visione semplicistica dell'AIgen come mero strumento di produttività. È un potente agente di cambiamento comportamentale, cognitivo e sociale. La sfida futura sarà quella di progettare e adottare modalità di interazione che promuovano un "aumento" delle capacità umane, mitigando i rischi di una "sostituzione" che, sebbene efficiente, potrebbe rivelarsi costosa per il benessere e l'autonomia individuale e collettiva (cfr. Laudadio, 2025).

Dai risultati dello studio, sembrerebbe emergere che il riporto al dominio e all'ambito, sembra nascondere comportamenti che singolarmente invece potrebbero assumere una certa significatività. Ad esempio, la lettura si sposta da libri a post e cresce il consumo di podcast, in un quadro di sostituzione più che di riduzione o aumento.

Sulla base dei risultati emersi e dei gap identificati nella letteratura, possiamo proporre sei dimensioni strategiche per orientare la ricerca futura sugli effetti dell'AI Generativa:

1. Studi Longitudinali a Lungo Termine. È urgente condurre studi longitudinali che seguano coorti di utenti per periodi prolungati (12-24 mesi), al fine di distinguere gli effetti transitori di adattamento da quelli strutturali e potenzialmente irreversibili. Particolare attenzione dovrebbe essere rivolta all'evoluzione delle capacità di memoria di lavoro, attenzione sostenuta e pensiero critico nel tempo, utilizzando misure oggettive (es. test neuropsicologici) oltre a quelle auto-riferite.
2. Neuroscienze Cognitive e Plasticità Cerebrale. La ricerca futura dovrebbe integrare metodologie neuroscientifiche (fMRI, EEG) per indagare se e come l'uso intensivo di AIgen modifichi i pattern di attivazione cerebrale associati a compiti di problem-solving, memoria e creatività. Studi sulla plasticità cerebrale potrebbero rivelare se l'offloading cognitivo induca riorganizzazioni funzionali durature, analogamente a quanto osservato con l'uso intensivo di GPS sulla navigazione spaziale.

3. Differenze Individuali e Fattori di Protezione. È necessario identificare i fattori individuali (es. tratti di personalità, stili cognitivi, competenze metacognitive) che modulano la vulnerabilità o la resilienza agli effetti negativi. Comprendere perché alcuni individui mantengono autonomia e pensiero critico mentre altri sviluppano dipendenza potrebbe informare interventi educativi e strategie di utilizzo consapevole.
4. Impatto sullo Sviluppo in Età Evolutiva. Dato che le generazioni più giovani sono esposte all'AIGen in fasi critiche dello sviluppo cognitivo e sociale, è imperativo condurre ricerche su bambini e adolescenti. Quali sono gli effetti sull'acquisizione di competenze fondamentali come la scrittura, il ragionamento matematico e le abilità sociali? Esistono "finestre critiche" di sviluppo particolarmente vulnerabili?
5. Progettazione di Interfacce per l'*Augmentation Responsabile*. La ricerca dovrebbe collaborare con designer e sviluppatori per co-progettare interfacce che promuovano un uso dell'AI orientato all'aumento (augmentation) piuttosto che alla sostituzione. Ciò potrebbe includere meccanismi di *friction* strategica che incoraggino la riflessione prima dell'offloading, feedback trasparenti sul livello di dipendenza, e strumenti per il monitoraggio e la regolazione dell'uso.
6. Implicazioni Sociali, Etiche e Politiche. Infine, è necessario un dialogo interdisciplinare che coinvolga scienze sociali, etica, diritto e policy-making. Come dovrebbero essere regolamentati gli strumenti AIGen in contesti educativi e professionali? Quali competenze "umane" dovrebbero essere protette e coltivate come patrimonio irrinunciabile? Come garantire equità di accesso evitando che l'AIGen amplifichi divari esistenti?

In sintesi, la ricerca futura deve adottare un approccio sistematico e multi-livello, che integri neuroscienze, psicologia, sociologia e design, per comprendere e orientare la co-evoluzione tra esseri umani e intelligenza artificiale generativa in modo che sia sostenibile, equo e rispettoso della dignità e dell'autonomia umana.

## Bibliografia

- Abbas, N., Pickering, B., Atwell, E., & Walker, A. (2024). Exploring the double-edged sword: A mixed methods study of ChatGPT use and academic performance. *International Journal of Educational Technology in Higher Education*, 21, Article 52. <https://doi.org/10.1186/s41239-024-00487-7>
- Almoqbel, M. Y. (2024). Talking to machines: Personas and behavioral patterns in Gen AI interactions. *Journal of Posthumanism*, 4(3), 411-434. <https://doi.org/10.63332/joph.v4i3.3243>
- Alshakhs, S. (2025). Arabic validation of Large Language Model Dependency Scale (D12): A psychometric study. Preprint.
- Bai, H., Voelkel, J. G., Muldowney, S., Eichstaedt, J. C., & Willer, R. (2025). AI-generated messages can be used to persuade humans on policy issues. OSF Preprints. [https://doi.org/10.31219/osf.io/stakv\\_v6](https://doi.org/10.31219/osf.io/stakv_v6)
- Castiello, S., & Costa, M. A. (2025). Generative AI as digital therapy: What we know and how to make it better? OSF Preprints. [https://doi.org/10.31234/osf.io/b26zh\\_v1](https://doi.org/10.31234/osf.io/b26zh_v1)
- Costello, T. H., Pennycook, G., & Rand, D. G. (2024). Durably reducing conspiracy beliefs through dialogues with AI. *Science*, 385(6714), 1230-1235. <https://doi.org/10.1126/science.adq1814>

- Crawford, J., Cowling, M., Ashton-Hay, S., Kelder, J.-A., Middleton, R., & Wilson, K. (2024). When artificial intelligence substitutes humans in higher education: The cost of loneliness, student success, and retention. *Studies in Higher Education*. <https://doi.org/10.1080/03075079.2024.2326956>
- De Freitas, J., Matz, S. C., Friedman, S. A., & Kim, K. (2025). AI companions reduce loneliness. *Journal of Consumer Research*. <https://doi.org/10.1093/jcr/ucaf040>
- Duong, C. D., Vu, T. N., & Ngo, T. V. N. (2024). Applying the theory of psychological strain to explain compulsive ChatGPT use. *International Journal of Human-Computer Interaction*. <https://doi.org/10.1080/10447318.2024.2410777>
- Dvorak, F., Stumpf, R. P., Fehrler, S., & Schudy, S. (2024). Generative AI triggers welfare-reducing decisions in humans. *arXiv preprint*. <https://doi.org/10.48550/arxiv.2401.12773>
- Fang, C. M., Liu, A. R., Danry, V., Lee, E., Chan, S. W. T., Pataranutaporn, P., Maes, P., Phang, J., Lampe, M., Ahmad, L., & Agarwal, S. (2025). How AI and human behaviors shape psychosocial effects of chatbot use: A longitudinal randomized controlled study. *arXiv*. <https://arxiv.org/abs/2503.17473>
- Folk, D. P., & Dunn, E. W. (2025). Turning to AI for companionship predicts increased loneliness four months later. *OSF Preprints*. [https://doi.org/10.31234/osf.io/bq7v3\\_v1](https://doi.org/10.31234/osf.io/bq7v3_v1)
- Georgiou, G. P. (2025). ChatGPT produces more "lazy" thinkers: Evidence of cognitive engagement decline. *arXiv preprint*. <https://doi.org/10.48550/arxiv.2507.00181>
- Glickman, M., & Sharot, T. (2024). How human–AI feedback loops alter human perceptual, emotional and social judgements. *Nature Human Behaviour*, 8, 1888-1901. <https://doi.org/10.1038/s41562-024-02077-2>
- Goh, A. Y. H., Hartanto, A., & Majeed, N. M. (2025). Generative artificial intelligence dependency: Scale development, validation, and its motivational, behavioral, and psychological correlates. *OSF Preprints*. [https://doi.org/10.31234/osf.io/aphtb\\_v1](https://doi.org/10.31234/osf.io/aphtb_v1)
- Han, Z. (2025). Artificial intelligence and social interaction: Evidence from gift money expenditure. *Scientific and Social Research*, 7(8), 226-236. <https://doi.org/10.26689/ssr.v7i8.12018>
- Head, K. R. (2025). Minds in crisis: How the AI revolution is impacting mental health. *Journal of Mental Health and Clinical Psychology*, 9(2), Article 1352. <https://doi.org/10.29245/2578-2959/2025/2.1352>
- Hirabayashi, S., Jain, R., Jurković, N., & Wu, G. (2024, June 4). Harvard Undergraduate Survey on Generative AI (arXiv:2406.00833v1) [Preprint]. *arXiv*. <https://arxiv.org/abs/2406.00833>
- Jadhav, A. (2025). Distributed atrophy: The cognitive costs of AI offloading. *Cognitive Science Quarterly*, 15(2), 112-128.
- Jiang, Y., Brinkmann, L., Nußberger, A.-M., Soraperra, I., & Bonnefon, J.-F. (2025). Humans learn to prefer trustworthy AI over human partners. *arXiv preprint*. <https://doi.org/10.48550/arxiv.2507.13524>
- Kim, J., Lee, S., & Park, M. (2025). Longitudinal effects of social chatbot interaction on loneliness and social anxiety in university students. *Journal of Medical Internet Research*, 27(1), e52341. <https://www.jmir.org/2025/1/e65589>

- Kim, R. (2025). AI-mediated cognitive distortions and digital echo delusions: Toward a resonant amplification framework. OSF Preprints. [https://doi.org/10.31234/osf.io/jcagk\\_v1](https://doi.org/10.31234/osf.io/jcagk_v1)
- Kleinert, T., Waldschütz, M., Blau, J., Heinrichs, M., & Schiller, B. (2025, June 25). AI outperforms humans in establishing interpersonal closeness in emotionally engaging interactions – but only when labelled as human. Research Square. <https://doi.org/10.21203/rs.3.rs-6803722/v1>
- Klimova, B., & Pikhart, M. (2025). Exploring the effects of artificial intelligence on student and academic well-being in higher education: A mini-review. *Frontiers in Psychology*, 16, Article 1498132. <https://doi.org/10.3389/fpsyg.2025.1498132>
- Kościelniak, M., & Bielecki, M. (2024). The psychology of generative AI in higher education: Mapping benefits and risks. *Qeios*. <https://doi.org/10.32388/x1tss>
- Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. MIT Media Lab. <https://www.media.mit.edu/publications/your-brain-on-chatgpt/>
- Laudadio, A. (2025) IA, opportunità o minaccia? Due generazioni a confronto. Agenda Digitale. <https://www.agendadigitale.eu/cultura-digitale/ia-opportunita-o-minaccia-due-generazioni-a-confronto/>
- Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25). Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713778>
- Lin, X., & Tan, H. (2025). A systematic review of generative AI in K–12: Mapping goals, activities, roles, and outcomes via the 3P model. *Systems*, 13(10), Article 840. <https://doi.org/10.3390/systems13100840>
- Luo, X., Ghosh, S., Tilley, J. L., Besada, P., & Wang, J. (2025). "Shaping ChatGPT into my digital therapist": A thematic analysis of social media discourse on using generative artificial intelligence for mental health. *Digital Health*, 11. <https://doi.org/10.1177/20552076251351088>
- Maral, S., Nayci, N., Bilmez, H., Erdemir, E.I., Satici, S.A. (2025) Problematic ChatGPT Use Scale: AI-Human Collaboration or Unraveling the Dark Side of ChatGPT. *Int J Ment Health Addiction* (2025). <https://doi.org/10.1007/s11469-025-01509-y>
- Matueny, R. M., & Nyamai, J. J. (2025). Illusion of competence and skill degradation in artificial intelligence dependency among users. *International Journal of Research and Scientific Innovation*, 12(5), 1263-1275. <https://doi.org/10.51244/ijrsi.2025.120500163>
- Namvarpour, M., Roberts, S. G., & Perfors, A. (2025). Romance, relief, and regret: Teen narratives of chatbot overreliance on Character.AI. arXiv preprint. <https://doi.org/10.48550/arxiv.2507.15783>
- Papagianni, A., & Spatoula, V. (2025). Understanding Gen Z's motivations for AI chatbot use and their associations with loneliness and mental health. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 19(1), Article 4.

- Qiu, T. A., He, Z., Chugh, T., & Kleiman-Weiner, M. (2025). The Lock-in Hypothesis: Stagnation by Algorithm. arXiv. <https://arxiv.org/abs/2506.06166>
- Rohilla, A. (2025). Impact of excessive AI tool usage on the cognitive abilities of undergraduate students: A mixed method study. *Assaj International Journal*, 4(1), 115-128. <https://doi.org/10.55966/assaj.2025.4.1.0115>
- Skjuve, M., Følstad, A., & Brandtzæg, P. B. (2023). A longitudinal study of self-disclosure in human-chatbot relationships. *Interacting with Computers*, 35(1), 24–39. <https://doi.org/10.1093/iwc/iwad022>
- Thorp, S., Slupphaug, K., Rimol, L. M., Lervik, S., & Hoel, S. K. B. (2025). Conformity towards humans versus AI in different task domains: The type of task matters. *Journal of Cognitive Enhancement*, 9(1), 45-58. <https://doi.org/10.1080/29974100.2025.2540762>
- Williams-Ceci, S., Jakesch, M., Bhat, A., Kadoma, K., & Zalmanson, L. (2025). Biased AI writing assistants shift users' attitudes on societal issues. OSF Preprints. [https://doi.org/10.31234/osf.io/mhjn6\\_v2](https://doi.org/10.31234/osf.io/mhjn6_v2)
- Wojtowicz, Z., & DeDeo, S. (2024). AI makes thinking easy but cooperation hard. *Proceedings of the National Academy of Sciences*, 121(12), e2316809121.
- Yankouskaya, A., Sui, J., & Humphreys, G. W. (2025). Development and validation of the LLM Dependency Scale (LLM-D12). arXiv preprint.
- Yu, S. C., Chen, H. R., & Yang, Y. W. (2024). Development and validation the problematic ChatGPT Use Scale: A preliminary report. *Current Psychology*, 43(31), 26080–26092. <https://doi.org/10.1007/s12144-024-06259-z>
- Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, 11, Article 28. <https://doi.org/10.1186/s40561-024-00316-7>
- Zhang, H., Wang, Y., & Liu, C. (2024). Development and validation of the AI Chatbot Dependence Scale. *Computers in Human Behavior*, 152, Article 108087.
- Zhang, R. Z., Kyung, E., Longoni, C., Cian, L., & Mrkva, K. (2024). AI-induced indifference: Unfair AI reduces prosociality. *Cognition*, 245, Article 105937. <https://doi.org/10.1016/j.cognition.2024.105937>
- Zhu, F., & Zou, W. (2024). The role of generative AI in human creative processes: Experimental evidence. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4676053>

## **Guerra e media: come sinistra e destra raccontano i conflitti in modo diverso – di Edvige Danna, Docente a contratto Unicusano**

*La comunicazione dei conflitti varia secondo l'orientamento politico: la sinistra enfatizza vittime civili e diplomazia, la destra sicurezza e identità nazionale. I social media amplificano queste differenze, creando echo-chamber che polarizzano il dibattito pubblico contemporaneo*

In un'epoca in cui la comunicazione è ubiqua – dai media tradizionali ai social network – la narrazione del conflitto assume una dimensione strategica: non si tratta soltanto di riportare fatti, ma di costruire senso, identità, legittimità. La guerra diventa non solo un evento militare ma anche un frame comunicativo. In questo contesto, lungo l'asse politico che va dalla sinistra alla destra, emergono stili, priorità, metafore e simboli differenti. Questa analisi intende mettere a fuoco tali differenze, riflettendo sulle logiche che regolano il racconto della guerra, e sul ruolo del digitale nel riplasmarlo.

### **Concetti teorici: framing, agenda-setting, ideologia**

Per orientarsi è utile richiamare qualche concetto chiave della comunicazione politica.

Il concetto di *framing* si riferisce al modo in cui i media e gli attori comunicativi selezionano e enfatizzano alcuni aspetti della realtà, escludendone altri, così da costruire “cornici” interpretative.

La teoria dell'*agenda-setting* poi spiega come i media non diano risposte su cosa pensare, ma influenzino su cosa pensare, ossia quali temi assumono rilevanza pubblica.

Sul piano ideologico, studi recenti mostrano che l'orientamento politico (sinistra/destra) incide sulle valutazioni morali, ad esempio sulla tolleranza rispetto ai danni collaterali in guerra: uno studio ha evidenziato che “gli individui orientati a sinistra esprimono in genere un sostegno maggiore per la minimizzazione delle vittime civili rispetto a quelli orientati a destra”.

Questi strumenti teorici ci aiutano a comprendere come le narrazioni della guerra siano funzionali non solo a descrivere, ma anche a stabilire chi siano i protagonisti “buoni” o “cattivi”, quale posizione avere rispetto al conflitto, e quali valori difendere o mettere in discussione.

### **Il racconto della guerra “da sinistra”**

Quando partiamo da una prospettiva di sinistra, alcune caratteristiche ricorrono frequentemente nel racconto dei conflitti:

- Enfasi sul civile, sulla vittima, sulla diplomazia: la narrazione tende a concentrarsi sul tema della sofferenza delle popolazioni, sull'impatto umanitario, sulla dimensione internazionale, sulle logiche economiche che alimentano la guerra. Ad esempio, uno studio sulla copertura del

confitto tra Ucraina e Russia ha evidenziato che certi media hanno privilegiato il “frame dell’interesse umano” (human-interest) piuttosto che quello della legittimità militare.

- Critica alle logiche di potenza e imperialismo: la sinistra tende a interpretare la guerra come esito di relazioni asimmetriche, di competizione geopolitica, di interessi economici. In questo senso, il racconto non è neutrale ma porta con sé una visione critica dei poteri.
- Promozione della pace, cooperazione, diritti umani: il discorso si sposta spesso verso soluzioni alternative alla forza militare, verso l’interdipendenza, verso interventi internazionali umanitari.
- Lessico e simboli: parole come «invasione», «occupazione», «resistenza popolare», «diritti» sono frequenti; l’attenzione è rivolta anche alle vittime civili, ai rifugiati, agli effetti collaterali del conflitto.

Questo stile narrativo, pur variegato, tende ad associare la guerra a fallimento politico, a costo sociale, a fallimento della diplomazia. Si tende a cavalcare l’idea secondo cui si è necessariamente nel giusto, senza che vi sia possibilità di replica o di accusa. Si enfatizzano macrotematiche umanitarie e si fa leva sui sentimenti di *pietas* e compassione.

### **Il racconto della guerra “da destra”**

Sul versante opposto, la destra propone un racconto della guerra con altre coordinate comunicative:

- Sicurezza, identità, minaccia esterna: la guerra viene spesso inquadrata come uno scontro tra “noi” e “loro”, come difesa della nazione, della civiltà, dei valori occidentali. Il discorso privilegia la protezione, la coesione nazionale, la reazione a un’aggressione.
- Legittimazione dell’uso della forza, ruolo militare valorizzato: nella narrazione di destra, l’azione militare può essere descritta come necessaria, giustificata, magari addirittura come forma di intervento morale o strategico. Studi mostrano che gli orientamenti politici di destra tendono a mostrare una maggiore accettazione delle vittime civili collaterali in scenari di autodifesa.
- Lessico e simboli: presenza di parole come «difesa», «minaccia», «identità», «valori», «ordine». La guerra è raccontata anche come elemento di rigenerazione o affermazione nazionale.
- Relazione con i media e le “fake news”: spesso, questo racconto denuncia una presunta parzialità dei media mainstream (“liberal”, “di sinistra”) e valorizza fonti alternative, reticenze sulla verità ufficiale, retoriche di complotto o censura.

In sintesi, la destra narra la guerra con un taglio “essenziale”: chi siamo, chi ci attacca, perché ci difendiamo, quali sono i valori da salvaguardare. Racconta un pragmatismo atto alla risoluzione dei problemi che da sempre caratterizza il taglio narrativo dei discorsi politici che le sono propri.

### **Lo spazio digitale e l'amplificazione delle polarizzazioni**

Il passaggio dalle audience passive ai mezzi digitali ha radicalmente modificato il contesto comunicativo:

Le piattaforme social modificano la dinamica di visibilità: algoritmi privilegiano contenuti forti, emozionali, polarizzanti. Ad esempio, uno studio sul conflitto in Gaza ha mostrato come la ripetizione di pochi frame selezionati e l'interazione digitale possano rafforzare narrative unilaterali.

Il digitale consente l'instantaneità, la multimedialità e la partecipazione: video, immagini, meme, post diventano strumenti della narrazione bellica. Uno studio sulla copertura del conflitto Russia-Ucraina in media cinesi ha rilevato che, sulla piattaforma Douyin, i media mainstream usavano frames verbal-visivi molto marcati: “macchine di guerra”, “ufficiali”, “stress visivo/sonoro”.

Le comunità digitali rafforzano le identità ideologiche: l'asse sinistra/destra si amplifica con echo-chamber, bolle di filtro, contestazione delle fonti. I media o gli attori con orientamento definito trovano nella rete terreno fertile per costruire versioni alternative del racconto ufficiale. In questo contesto, non soltanto ciò che viene raccontato conta, ma come viene raccontato e da chi. L'orizzonte comunicativo si fa più fluido, ma anche più vulnerabile alle distorsioni e alle polarizzazioni.

Interessante anche constatare la nascita di contenuti strutturati per la *ridicolizzazione del conflitto*, sostenuto da una parte politica piuttosto che un'altra: immagini sarcastiche, video costruiti con intelligenza artificiale il cui fine ultimo è sostenere esclusivamente le proprie idee e non un bene più alto che andrebbe nella direzione di sostegno ai popoli in guerra.

### **Analisi comparata: alcune evidenze empiriche**

Di seguito si presentano alcune evidenze significative utili a rafforzare l'analisi.

Uno studio su 417 articoli di The Guardian e China Daily (17 febbraio-16 marzo 2022) ha rilevato che il Guardian ha usato prevalentemente il frame “interesse umano”, mentre China Daily ha privilegiato il frame «responsabilità» (“attribution of responsibility”).

Un'analisi comparativa di stampa statunitense, europea e latino-americana durante la guerra in Iraq (2003) ha mostrato che i media statunitensi citavano quasi esclusivamente fonti della coalizione, offrendo un quadro più ristretto; i media europei e latino-americani usavano una gamma più ampia di fonti e prospettive.

In uno studio sui media della Taiwan riguardo al conflitto Palestina-Israele, è emersa che testate conservative/pro-Cina tendevano a un racconto pro-Palestina, focalizzandosi sull'impatto umanitario israeliano, mentre testate progressiste/pro-Taiwan puntavano su un racconto pro-Israele.

Nel contesto degli orientamenti ideologici individuali, uno studio recente su spargimento di vittime civili in guerre ha evidenziato che l'essere politicamente orientati "a sinistra" associa una maggiore rigidità nel rifiuto delle vittime civili rispetto a chi è orientato "a destra".

Tali evidenze confermano come la narrazione della guerra non sia neutra: dipende dal media system, dalla geografia, dall'ideologia, dal controllo delle fonti, dal grado di digitalizzazione.

### **Perché queste differenze comunicative contano**

Le differenze tra il racconto della guerra da sinistra e da destra non sono semplicemente un gioco stilistico, ma implicano conseguenze concrete.

Nella formazione dell'opinione pubblica, i frame selezionati orientano non solo la percezione del conflitto, ma anche cosa pensiamo sia giusto fare (es. intervenire, armare, dialogare).

Inoltre, entra in gioco anche il concetto di *legittimazione politica*, in quanto la narrazione sostiene (o delegittima) le decisioni di politica estera, di uso della forza, di intervento internazionale. L'escamotage più semplice e impattante a livello emotivo è delegittimare le idee e le azioni avversarie additandole facendo riferimenti a fatti storici ed eventi catastrofici, si pensi alle classiche frasi: «siete fascisti», «siete dei poveri comunisti», entrati nel linguaggio comune fino ad essere svuotati di senso e valore.

Si tratta di una costante polarizzazione che alimenti rotture e scontri interni; quando la comunicazione si fa ideologica, la società si divide non solo sui fatti della guerra, ma su valori fondamentali (diritti civili, identità, alleanze) e spesso segue l'onda del proprio partito politico senza guardare complessivamente e con distacco alle evidenze fattuali.

La conseguenza è una generale crisi di fiducia nei confronti dei media poiché, in un contesto in cui ogni racconto è politicamente segnato, la credibilità dei media è messa alla prova. Si moltiplicano le accuse di "propaganda" e aumentano fenomeni di travisamento della realtà. La digitalizzazione è responsabile di questa amplificazione e la diffusione di narrazioni polarizzate sui social può generare radicalizzazioni rapide, camere dell'eco e difficoltà di dialogo e sano confronto pubblico. In tal modo, il racconto della guerra diventa anche un racconto interno alle società, che influenza la coesione, la democrazia, la capacità di elaborare politiche estere responsabili.

### **Verso un'ecologia del racconto: spunti per il futuro**

Alla luce delle analisi, si può suggerire qualche orientamento operativo per "governare" meglio la comunicazione della guerra.

In primis è necessario promuovere *pluralismo nelle fonti*: una copertura che includa voci diverse, fonti marginalizzate, prospettive umanitarie e strategiche contemporaneamente. Come ricordato in Francesco Olmastroni (2014), un modello «ciclico» tra opinione pubblica, media ed élite permette di evitare l'assunzione passiva delle cornici dominanti.

I frame utilizzati nella costruzione dei discorsi politici devono essere trasparenti, tant'è vero che i media dovrebbero chiarire che la scelta del frame implica una certa visione di mondo, non limitandosi a riportare solo ciò che è accaduto.

L'educazione mediatica deve essere alla base dello sviluppo di responsabilità e consapevolezza, i cittadini vanno resi consapevoli non solo dei contenuti, ma delle *cornici comunicative* che li veicolano, soprattutto nella dimensione digitale.

Sviluppare un *dialogo inter-ideologico* è poi un altro tassello che completerebbe il quadro di una comunicazione più efficace; la distinzione tra sinistra e destra presuppone che i due campi si parlino, se invece si comunicano solo verso i propri "fan", la polarizzazione diventa endemica.

Infine, le piattaforme social devono chiarire il loro ruolo nell'amplificazione delle narrazioni belliche e adottare policy che mitighino la propagazione incontrollata di messaggi polarizzanti.

Analizzare il racconto della guerra da sinistra a destra significa riconoscerne la *non neutralità*: ogni comunicazione porta con sé valori, priorità, scelte. In un'epoca in cui la velocità dell'informazione e la potenza degli algoritmi digitali amplificano le differenze, è essenziale tornare a una riflessione sulla forma e sul contenuto delle narrazioni belliche. Non si tratta di decidere "chi ha ragione", ma di rendere visibili i meccanismi attraverso cui le opinioni si costruiscono, si polarizzano e influenzano le decisioni pubbliche.

Abbracciare un racconto piuttosto che un altro significa non solo banalizzare il problema, ma non comprenderlo. Osservare i fenomeni da un'unica prospettiva, pensando di essere dalla parte giusta della storia, minimizza la complessità delle decisioni geopolitiche che stanno alla base della costruzione di questo nuovo mondo. Occorre pertanto mantenere lucidità, informarsi da più fonti, anche quelle ritenute scomode o non in linea con le proprie convinzioni politiche, poiché si evincerà quanto profondi e differenti siano i dettagli presi in considerazione. Solo così potremo sperare in un panorama comunicativo capace non solo di informare, ma di aprire spazi di comprensione e non di conflitto interno.

## Bibliografia

- Benanti Paolo (2024). *Il crollo di Babele. Che fare dopo la fine del sogno di Internet?* San Paolo Edizioni.
- Lakoff G., Brindisi D. (2020). *Non pensare all'elefante! Come riprendersi il discorso politico. Le tecniche per battere la destra e reinventare la sinistra, a partire dalle parole che usiamo ogni giorno.* Chiarelettere
- Luna Riccardo (2025). *Qualcosa è andato storto. Come i social network e l'intelligenza artificiale ci hanno rubato il futuro.* Solferino.
- Mieli Paolo (2025). *Il prezzo della pace. Quando finisce una guerra.* Rizzoli.
- Elad-Strenger, J., Statman, D., Kessler, T. (2025). *Left-Right Ideological Differences in Moral Judgments: The Case of Acceptance of Collateral Civilian Killings in War.* European Journal of Social Psychology, 55(4), 565-588.
- Hermilania, H. (2025). *The Power of Framing: Examining BBC's Representation of the Ukraine War in March 2022.* Anaphora: Journal of Language, Literary, and Cultural Studies, 7(2).

Chai, M. H., Liu, Y. H. (2024). *Framing Analysis of the Conflict between Russia and Ukraine* in China Daily and The Guardian. Quantum Journal of Social Sciences and Humanities, 6(4).

Olmastroni, F. (2014). *Framing War: Public Opinion and Decision-Making in Comparative Perspective*. Routledge.

Wang, Y., Sannusi, S. N., Abd Kadir, S. (2024). *News Framing of the 2022 Russian-Ukrainian Conflict: A Comparative Analysis of CNN (USA) and CCTV (China) from a War and Peace Journalism Perspective*. e-Bangi: Journal of Social Sciences & Humanities.

Almahallawi, W., Awais, I. A. (2025). *Framing News in the Time of War: A Comparative Study between International and Regional Televisions*. Journal of Media and Communication Research.

## **Marketing, Intelligenza Artificiale e creatività: uno spartiacque tecnologico – di Massimiliano Moschin, Dottore di ricerca in sociologia e internet studies, IUSVE**

*La tecnologia offre al marketing possibilità enormi, ma rischia di appiattire linguaggi e creatività. A fare la differenza non saranno gli strumenti, bensì cultura, qualità degli input e capacità dei professionisti di guidare l'intelligenza artificiale invece di farsene guidare*

Il marketing è oramai una disciplina affermata con più di 60 anni di storia accademica a partire dai primi studi pionieristici di McCarthy e Kotler, anche se a ben vedere le sue origini sono da ricercare ben più in là nel tempo (Earle, 2009), praticamente da quanto esiste l'interazione umana e lo scambio di merci in qualche modo la strategia di valorizzazione di un bene in una dinamica di scambio è entrata a far parte del nostro DNA. Se quindi è qualcosa che ci portiamo dentro, mai come in tempi recenti questo ambito sta subendo trasformazioni a una velocità sempre più crescente (Bist, Agarwal, Aini, Khofifah, 2022), per cui serve ripensare il rapporto tra mente umana e cibernetica, il rischio infatti è che senza fondamenta culturali solide, la nostra “torre del sapere” crolli o peggio, diventi una sorta di Babele al contrario, dove gli unici che ne comprendono il linguaggio sono gli abitanti reclusi nell’edificio, mancando la volontà di comprendere il mondo oltre alla limitazione di un muro virtuale.

### **Il pensiero creativo imbrigliato in un recinto digitale**

Il senso di questa metafora è spiegabile attraverso un paradosso in cui chi si occupa di marketing spesso ricade. La tecnologia digitale dal punto di vista delle potenzialità che assicura a livello di conoscenza di un vasto patrimonio informativo e realizzazione operativa dovrebbe rappresentare sulla “carta” (mi si perdoni l’allusione agli old media) un enorme strumento accrescitivo di idee ed espressività (Shugan, 2004). Nella pratica questo si verifica solo in parte, dato che l'eccessiva praticità delle soluzioni che offre spesso diventa una gabbia creativa (Vakratsas, Wang, 2020), facendo ricadere l'espressione in linguaggi molto simili, su cui il modellamento di partenza di ambienti e piattaforme digitali ha un ruolo non secondario (Hearn, 2020), che si esprimono attraverso immagini e aspetti visivi stereotipati (Runco, 2015).

La digitalizzazione è un’opportunità potente per velocizzare e migliorare i processi, ma se dialogare con un algoritmo, eseguire query semantiche verso un database o impiegare modelli di Intelligenza Artificiale risultano esperienze molto efficienti nel contesto di compiti ripetitivi e standardizzabili, sono un po’ meno efficaci nel riconoscere la conoscenza umana e il funzionamento della sua creatività (Holford, 2019). Qualche esempio? Il trend di utilizzare immagini e video stock da banche dati, magari con contenuti accessibili a livello gratuito o in licenza creative commons, ha standardizzato talmente tanto la pubblicità online e a cascata quella offline, che sono nati addirittura fandom dedicati a [una modella che appare in una percentuale altissima di queste fotografie](#) di uno dei più celebri fornitori di contenuti (Ramish, Bashir, Jalees, 2019). Se quindi buona parte della creatività parte dagli stessi spunti visuali, alla lunga il linguaggio rischia di appiattirsi ed essere autoreferenziale,

ecco quindi la nostra baba al contrario erigersi anche sul pensiero, con un recinto creativo auto-imposto.

### **Opportunità tecnologiche: Intelligenza Artificiale**

In questo modo se l'utilizzo della tecnologia fino adesso per certi versi è stato un'arma a doppio taglio per la comunicazione e il marketing, l'Intelligenza Artificiale generativa, la grande novità che ha sconvolto il mondo della comunicazione negli ultimi anni, rischia di essere un altro strumento controverso (Grewal, Satornino, Davenport, Guha, 2025). Il successo dell'adozione lo testimonia, visto che in Italia al 2024 più dell'8% delle aziende ha adottato sistemi di Intelligenza Artificiale e in Europa più del 13% (Fonte: Istat), mentre le persone utilizzano sempre di più tecnologie come strumenti generalisti, come il noto ChatGPT di OpenAI per questioni personali da consigli medici o psicologici (Savoldi e altri 2025), o quelli verticali come Midjourney per generare contenuti visuali a discapito dei precedenti strumenti. Se però tanto è semplice creare contenuti con questi strumenti attraverso la formulazione di prompt efficaci è altrettanto facile cadere nei cliché espressivi, basti pensare ai vari trend derivati dal primo ciclo di video non-sense di "SkibidiBoppy" con l'uscita di Veo3 di Google, strumento per la generazione di video. La bacheche dei nostri profili social si sono riempite di filmati che, nella loro non linearità semantica, seguono lo stesso schema narrativo con personaggi abbastanza ricorrenti. La morale? E' diventata una forma di linguaggio, cosa non infrequente online, ma a son di replicarlo l'effetto saturazione è dietro l'angolo (Atienza-Barthelemy, Losada, Benito, 2025).

Fin qui nulla di terribile, il problema è che se tutti iniziano a usare gli stessi strumenti, con risultati uniformi, la comunicazione tenderà ad appiattarsi e omologarsi, perdendo di efficacia. Se infatti anche nel contesto digitale la creatività risulta un elemento ancora determinante (Boienko, Yanchuk, Fedchenko, 2025), questa prassi si scontra con un rigido approccio performativo, basato più sull'efficienza di un risultato a breve, piuttosto che anche orientato all'efficacia di produzione di valore a lungo termine (Lies, 2021), soprattutto laddove l'automazione standardizza a discapito di approcci più agili e creativi, ma soprattutto quando la valutazione della creatività stessa parte da un punto di vista analitico di una macchina. Sia chiaro che la questione non è assolutamente la tecnologia, ma il suo utilizzo e soprattutto la conoscenza da cui attinge. Se infatti per produrre contenuti il bagaglio di conoscenza è il patrimonio visuale riversato online in questi anni, che abbiamo visto non brillare per forza per unicità, i risultati espressivi non potranno che essere piuttosto standardizzati. Soprattutto, forse l'aspetto più preoccupante, è che se deleghiamo le idee e il loro giudizio alla macchina oltre all'esecuzione, le soluzioni non potranno che essere rivisitazioni, anche bellissime, di cose già fatte. Come si può superare questo impasse? Con la qualità e la modalità di input che viene dato all'Intelligenza Artificiale, elemento su cui forse la mente umana può ancora superare quella digitale.

### **Il settore in Italia: perché è necessario usare bene la AI**

Guardando al settore della comunicazione e del marketing in Italia appare evidente come i professionisti non possano permettersi di subire l'Intelligenza Artificiale, ma debbano cavalcarne in prima persona l'onda. La motivazione è insita nella struttura imprenditoriale dell'economia italiana, siamo infatti il Paese delle PMI e questa conformazione di piccole e medie imprese, con pochi addetti e basso livello di innovazione (Moncada-Paternò-Castello, Ciupagea, Piccaluga, 2006), si riverbera anche sul comparto delle aziende della comunicazione e del marketing, che assommando a sé le diverse vocazioni più o meno digitali, generano un settore di più di 40.000 aziende, di cui la maggior parte è costituito da massimo tre dipendenti, con un livello di fatturato che per la metà si attesta entro i 100.000€ (Fonte: Camere di Commercio).

Si tratta comunque di un mercato, quello della comunicazione che nel paese vale più di 11 miliardi di investimenti con un trend positivo anno su anno e una quota di investimenti nel digitale che ormai vale più del 50% (Dati Osservatorio Internet Media del Politecnico di Milano). Il tema quindi è quanto di questo valore potenziale può essere assorbito da strumenti che integrino intelligenza artificiale e che impatto questa può avere sulla qualità delle campagne di marketing in termini di creatività? Il legame fra questi numeri e le potenzialità e rischi offerti dalla tecnologia si scontrano con il potenziale investimento che le imprese possono mettere in campo. Se le aziende top si appoggiano ad agenzie di alto livello, spesso multinazionali con piani media complessi e risorse significative, significa che tutto il grande comparto di imprese più piccole si affida ad agenzie di minor cubatura, ma non per forza con idee non altrettanto valide. Il problema con cui i più piccoli si scontrano però sono i budget. La tentazione quindi per le imprese, in costante necessità di ottimizzare i costi, è quella di far da sé attraverso strumenti che oggi, grazie all'interfaccia uomo macchina semplificata, sono estremamente intuitivi per generare prodotti credibili, un po' meno per idee davvero innovative (Ameen, Sharma, Tarba, Rao, Chopra, 2022).

### **Il ruolo dei dati: il cuore pulsante dell'Intelligenza Artificiale**

Infatti ci si sta concentrando sempre di più sulla semplicità e immediatezza, attraverso il prompt design e la crescente integrazione con il linguaggio naturale degli strumenti e sull'accuratezza visiva del risultato, a discapito di pensare al motore principale su cui la sofisticata tecnologia dell'Intelligenza Artificiale si basa, ovvero i dati (Brodie, 1989). Se i processi infatti possono essere resi estremamente veloci e precisi quello che fa ancora la differenza è il database di partenza e la richiesta che si fa interrogando lo strumento che deve analizzare questi dati. Su questo terreno la creatività può ancora giocare le sue carte, non solo continuando a cercare nuove strade, ma migliorando al massimo il proprio strumento di orientamento: le informazioni, considerando anche sistemi di AI proprietari e al lavoro su database chiusi o non accessibili all'esterno dell'organizzazione che li gestisce.

Continuare a impiegare quasi esclusivamente soluzioni accessibili sul mercato, senza un minimo di personalizzazione nell'uso e accesso ai dati e, soprattutto, senza considerare i rischi di esporre il proprio patrimonio informativo alla condivisione con soggetti terzi (Cadelon, Reeves, 2022), resta un valido deterrente per usare in modo scriteriato questa potente tecnologia. Resta chiaro però che ci troviamo davanti a una rivoluzione inevitabile, che non andrà subita ma affrontata consapevolmente,

sempre di più il futuro dell'attività umana, anche a livello di marketing, viaggerà attraverso l'innovazione tecnologica, ecco quindi che vanno considerate tutte le implicazioni dell'utilizzo di questi sistemi (Elmore, Lee, 2021), non esclusivamente per tagliare costi o tempo, ma per potenziare il pensiero umano anche in campo creativo (Nyholm, 2024). In questo senso non bisogna interrogarsi se e quando introdurre l'intelligenza artificiale in ambito marketing, ma come e secondo quali principi, per generare valore concreto.

## Conclusioni

In sintesi l'Intelligenza Artificiale quindi non va né demonizzata ma neppure presa come panacea e usata in modo acritico, è necessario per tutti i professionisti imparare a padroneggiarla ma soprattutto a utilizzarla in parallelo alle proprie modalità di lavoro, in modo che non sostituisca, ma potenzi le mansioni che fanno le persone. Se paradossalmente l'AI abbasserà gli sforzi in ambito di produzione, non è automatico che lo debba fare a livello di pensiero, tanto che qualcuno ipotizza che le agenzie e le divisioni di marketing potrebbero proprio progredire dal fare al pensare (Wahid, Mero, Ritala, 2025). Se questo aspetto sembra così disruptive, in realtà potrebbe essere un ritorno all'ideare soluzioni creative piuttosto che affrontare bisogni di problem solving più tattici che strategici, quando grazie alla tecnologia si potrebbe non aver bisogno di cercare i problemi, ma "solo" di trovare a essi soluzioni.

Nel pensare e nell'esprimere creatività sarà ancora possibile differenziare i professionisti dagli "amatori", soprattutto a livello di cultura creativa necessaria a generare grandi idee, anche con budget relativamente piccoli. Altrimenti se lo scontro si limiterà a scrivere il miglior prompt, legittimamente, qualcun altro userà l'intelligenza artificiale per soppiantare a contenuti banali e standardizzati, accontentandosi e non trovando necessità a rivolgersi ad esperti. La storia della pubblicità e del marketing ci insegna che una comunicazione efficace, grazie a una grande idea creativa, ha il potere di durare nel tempo spesso più a lungo dei brand e dei prodotti stessi che rappresenta. Per cui, largo alla tecnologia se potenzia le idee, invece di soppiantarle. Dato quindi che è nel DNA di noi creativi il marketing, ricordiamoci che dobbiamo andare oltre lo standard, altrimenti ci ritroveremo chiusi in una torre bellissima, ma che scricchiola.

## Bibliografia

Ameen, N., Sharma, G. D., Tarba, S., Rao, A., & Chopra, R. (2022). Toward advancing theory on creativity in marketing and artificial intelligence. *Psychology & marketing*, 39(9), 1802-1825.

Atienza-Barthelemy, J., Losada, J. C., & Benito, R. M. (2025). Modeling Information Diffusion on Social Media: The Role of the Saturation Effect. *Mathematics*, 13(6), 963.

Bist, A. S., Agarwal, V., Aini, Q., & Khofifah, N. (2022). Managing Digital Transformation in Marketing: "Fusion of Traditional Marketing and Digital Marketing". *International Transactions on Artificial Intelligence*, 1(1), 18-27.

- Boienko, O., Yanchuk, T., & Fedchenko, O. (2025). Creativity in digital marketing. *Three Seas Economic Journal*, 6(1), 19-26.
- Brodie, M. L. (1989). Future intelligent information systems: AI and database technologies working together. In *Readings in artificial intelligence and databases* (pp. 623-641). Morgan Kaufmann.
- Candelon, F., & Reeves, M. (Eds.). (2022). *The rise of AI-powered companies*. Walter de Gruyter GmbH & Co KG.
- Earle, T. (2009). Exchange systems in prehistory. In *Trade and exchange: archaeological studies from history and prehistory* (pp. 205-217). New York, NY: Springer New York.
- Elmore, J. G., & Lee, C. I. (2021). Data quality, data sharing, and moving artificial intelligence forward. *JAMA network open*, 4(8), e2119345-e2119345.
- Grewal, D., Satornino, C. B., Davenport, T., & Guha, A. (2025). How generative AI Is shaping the future of marketing. *Journal of the Academy of Marketing Science*, 53(3), 702-722.
- Hearn, G. (2020). The future of creative work: Creativity and digital disruption. In *The future of creative work* (pp. 1-12). Edward Elgar Publishing.
- Holford, W. D. (2019). The future of human creative knowledge work within the digital economy. *Futures*, 105, 143-154.
- Kotler, P. (1965). Behavioral models for analyzing buyers. *Journal of Marketing*, 29(4), 37-45.
- Lies, J. (2021). Digital marketing: Incompatibilities between performance marketing and marketing creativity. *Journal of Digital & Social Media Marketing*, 8(4), 376-386.
- McCarthy, E. J. (1965). E. Jerome McCarthy. *The Concept of the Marketing Mix*.
- Moncada-Paternò-Castello, P., Ciupagea, C., & Piccaluga, A. (2006). L'innovazione industriale in Italia: persiste il modello "senza ricerca"??. *L'industria*, 27(3), 533-552.
- Nyholm, S. (2024). Artificial intelligence and human enhancement: can AI technologies make us more (artificially) intelligent?. *Cambridge Quarterly of Healthcare Ethics*, 33(1), 76-88.
- Ramish, M. S., Bashir, A., & Jalees, T. (2019). Visual appeal of stock photographs affecting the consumers' attitude towards advertising. *GMJACS*, 9(1), 15-15.
- Runco, M. (2015). The Real Creativity Crisis. *Creativity & Human Development*.
- Savoldi, B., Attanasio, G., Gorodetskaya, O., Manerba, M. M., Bassignana, E., Casola, S., Negri, M., Caselli, T., Bentivogli, L., Ramponi, A., Muti, A., Balbo N. & Nozza, D. (2025). Generative AI Practices, Literacy, and Divides: An Empirical Analysis in the Italian Context. *arXiv preprint arXiv:2512.03671*.
- Shugan, S. M. (2004). The impact of advancing technology on marketing and academic research. *Marketing Science*, 23(4), 469-475.
- Vakratsas, D., & Wang, X. (2020). Artificial intelligence in advertising creativity. *Journal of Advertising*, 50(1), 39-51.
- Wahid, R., Mero, J., & Ritala, P. (2025). Technology-enabled democratization: Impact of generative AI on content marketing agencies. *Industrial Marketing Management*, 131, 1-16.



## **Serious game e AI: il futuro dell'apprendimento è immersivo e personalizzato - di Luca Cavone, Università di Pavia**

*I Serious Game offrono esperienze interattive basate sul gioco. L'intelligenza artificiale, integrata in questi strumenti, ne aumenta l'efficacia, adattando contenuti e difficoltà in tempo reale. Una combinazione che ridisegna i modelli di formazione*

L'era digitale, così come altri in settori, sta rivoluzionando anche il panorama della formazione, rendendo sempre più obsolete le metodologie tradizionali, basate sulla mera trasmissione di informazioni secondo un approccio prevalentemente "monodirezionale". In un contesto come quello attuale ma soprattutto futuro, caratterizzato da una continua evoluzione, l'esigenza di nuovi strumenti di apprendimento più efficace, coinvolgenti e, soprattutto, personalizzabili è diventata cruciale. La formazione deve quindi adattarsi di pari passo per rispondere alle esigenze individuali e alle sfide complesse del mondo moderno, dove l'apprendimento continuo è diventato fondamentale per rimanere competitivi.

In questo scenario in continua evoluzione, due fattori chiave stanno ridefinendo quella che sarà in futuro l'esperienza formativa: i Serious Game e l'Intelligenza Artificiale (AI). I Serious Game, approcci con finalità educative e formative che si avvalgono di meccanismi ludici, hanno già dimostrato una notevole capacità di coinvolgere i partecipanti, trasformando l'apprendimento in un'esperienza immersiva e interattiva. Contemporaneamente, l'Intelligenza Artificiale, nella sua rapida evoluzione, sta raggiungendo livelli sempre più spinti di sofisticazione, quali la capacità di analizzare grandi quantità di dati e prendere decisioni autonome (Agentic AI), diventando un fattore abilitante per la personalizzazione su ampia scala. L'integrazione di queste due dimensioni rappresenta quindi il prossimo passo evolutivo di una nuova modalità di formazione digitale. Non si tratta semplicemente di rendere l'apprendimento più interattivo, ma di sfruttare l'AI per adattare dinamicamente il Serious Game ai ritmi, agli stili e ai livelli di conoscenza dei singoli individui. Inoltre, la combinazione di Serious Game e AI offrirà la possibilità di creare percorsi formativi unici, in grado di adattarsi in tempo reale alle performance dell'utente, fornendo feedback mirati e contenuti più appropriati.

Questo articolo analizza quindi come questa sinergia non stia solo migliorando l'immersività e l'engagement, ma soprattutto come stia rendendo l'apprendimento profondamente personalizzato, delineando il futuro della formazione nell'era digitale.

### **Serious Game: dal gioco all'immersione educativa**

L'evoluzione della tecnologia ha permesso ai **Serious Game** di superare la mera funzione di intrattenimento per affermarsi come strumenti didattici evoluti (Prensky, 2001; Michael & Chen, 2006). A differenza dei metodi di formazione tradizionali, che spesso si basano su un approccio passivo (come lezioni frontali o la lettura di manuali), i Serious Game sfruttano i principi della **ludicizzazione** per stimolare il coinvolgimento attivo del discente (Deterding et al., 2011). L'obiettivo non è solo quello

educativo e formativo, ma diventa anche divertire e creare coinvolgimento, grazie all'integrazione di meccanismi di gioco – come sfide, obiettivi, ricompense e narrazione.

Il principale impatto dei Serious Game infatti risiede nella loro capacità di creare esperienze di **apprendimento immersive**, trasportando i partecipanti in ambienti simulati e realistici, che permettono loro di affrontare situazioni complesse e di applicare le conoscenze acquisite in un contesto ludico-pratico (Gee, 2007). Questo approccio, basato sul principio dell'**apprendimento esperienziale** ("learning by doing"), si è dimostrato significativamente efficace nel migliorare lo sviluppo di competenze pratiche e la capitalizzazione di concetti e informazioni (Lave & Wenger, 1991; Ke, 2009). Diversi settori utilizzano già i Serious Game per simulare scenari critici in un ambiente sicuro (es. Gestione aziendale, Medicina, etc...), dove gli errori non hanno conseguenze nel mondo reale ma diventano preziose opportunità di apprendimento continuo.

Come già riportato, i vantaggi tradizionali dei Serious Game includono l'incremento dell'**engagement** e della **motivazione** del discente (Deterding et al., 2011), in aggiunta a questi, la possibilità di avere un feedback immediato e concreto, fornito dalla dinamica del gioco, permette ai partecipanti di comprendere rapidamente le conseguenze delle loro azioni e nel caso correggerle in tempo reale, favorendo un circolo virtuoso di pratica e miglioramento continuo (Anderson & Schunn, 2000). Questa capacità di rendere l'apprendimento un'esperienza dinamica ha già dimostrato un grande potenziale per trasformare l'educazione, ponendo anche le basi per nuove esperienze basate sull'integrazione con tecnologie ancora più avanzate, quali ad esempio l'Intelligenza Artificiale.

### L'Intelligenza Artificiale come motore di personalizzazione

L'Intelligenza Artificiale (AI), nel suo range di applicazione sempre più ampio, rappresenta un punto di svolta significativo anche nell'evoluzione della formazione digitale, fungendo da catalizzatore per la **personalizzazione dell'apprendimento** e dell'esperienza formativa (Chen & Wang, 2020; Hwang & Huang, 2018). A differenza degli approcci formativi tradizionali che in passato e ancora oggi offrono percorsi prevalentemente standardizzati, l'AI consente di superare e andare oltre questa rigidità, rispondendo in modo mirato alle esigenze uniche di ogni discente.

Come noto, il ruolo cruciale dell'AI risiede nella sua capacità di elaborare e interpretare grandi quantità di dati e nel caso specifico **dati legati all'apprendimento** (Siemens, 2013; Baker & Inventado, 2014). Attraverso algoritmi specifici, l'AI oltre a registrare i progressi del percorso di apprendimento, può analizzare in dettaglio come questi vengono determinati, identificando lacune specifiche nella conoscenza, schemi di errori ricorrenti e, persino, i livelli di coinvolgimento emotivo. Questo tracciamento dettagliato permetterebbe una comprensione granulare del comportamento dell'utente lungo tutte le fasi del ciclo di apprendimento, superando la tradizionale valutazione basata solo sui risultati finali.

L'AI diventerebbe quindi un pilastro portante per i **sistemi di apprendimento adattivo**, che farebbero leva sui dati raccolti per ridefinire e modellare in modo mirato l'esperienza formativa del discente in tempo reale (Brusilovsky, 1996; Woolf, 2009). Ad esempio, un algoritmo potrebbe riconoscere che un discente sta incontrando difficoltà con un concetto specifico e, immediatamente, reindirizzarlo verso moduli didattici più semplici o fornire risorse aggiuntive per aiutarlo nella comprensione. Al contrario, se il discente mostra una rapida padronanza di un determinato argomento, l'AI potrebbe stimolare

ulteriormente la sua velocità di apprendimento, accelerando il percorso e proponendo contenuti più avanzati e mirati.

In questo modo la formazione diventa un'esperienza dinamica e individualizzata che si adatta costantemente, garantendo che ogni studente riceva il contenuto corretto e appropriato per il proprio livello di apprendimento. Questo fattore rappresenta una base per l'integrazione con i **Serious Game**, trasformando semplici esperienze immersive in potenti strumenti di crescita personalizzata (Gomes & Lopes, 2021; Nacke, 2013).

### **Serious Game e AI per l'apprendimento adattivo**

Abbiamo appena visto come l'ambito della formazione potrebbe beneficiare di una profonda rivoluzione trasformativa, basata sulla convergenza tra **Serious Game e Intelligenza Artificiale** (Gomes & Lopes, 2021; Nacke, 2013). L'integrazione dell'AI non si limiterebbe a potenziare le capacità tecniche dei giochi, ma li trasformerebbe in piattaforme di apprendimento intelligenti e dinamiche, capaci di agire come un "tutor" virtuale all'interno dell'ambiente di gioco (Woolf, 2009).

Il meccanismo chiave di questa convergenza è l'**apprendimento adattivo** (Brusilovsky, 1996), grazie all'analisi costante delle interazioni del giocatore, il suo ritmo di apprendimento e i suoi progressi (Baker & Inventado, 2014). Sulla base di questi dati, l'Intelligenza Artificiale modifica in tempo reale la difficoltà, il contenuto e la sequenza delle sfide proposte dal gioco (Hwang & Huang, 2018; Chen & Wang, 2020).

Per fare un esempio pratico, è possibile fare riferimento al concetto di difficoltà dinamica: se un discente dimostra di padroneggiare rapidamente un concetto, l'AI potrebbe introdurre scenari più complessi e stimolanti per mantenere alta la motivazione; al contrario, se incontra difficoltà, l'AI potrebbe semplificare le sfide o offrire esercizi di rinforzo mirati, per garantire la possibilità di raggiungere gli obiettivi iniziali. Questo assicura che il discente non sia mai né annoiato né sopraffatto, ma che continui l'esperienza di apprendimento in modo proficuo per quello che è il suo livello specifico.

In aggiunta a quanto già riportato, l'AI permetterebbe di ricevere un **feedback personalizzato e contestuale** (Anderson & Schunn, 2000). Invece di messaggi generici, il sistema può spiegare esattamente perché un errore è stato commesso all'interno di una dinamica di gioco, offrendo suggerimenti specifici per migliorare le performance. Questo approccio non solo ottimizza il tempo di apprendimento, ma rafforza anche l'immersione, poiché il gioco reagisce in modo intelligente e significativamente mirato alle azioni del singolo giocatore. La sinergia quindi tra l'immersione del Serious Game e l'intelligenza dell'AI crea un ambiente dove l'apprendimento è fluido, continuo e costantemente ottimizzato per le esigenze individuali.

### **Benefici, sfide e prospettive future dell'apprendimento immersivo e personalizzato**

È stato descritto come l'integrazione di Serious Game e Intelligenza Artificiale possa offrire vantaggi concreti che vanno oltre il semplice engagement, impattando direttamente sull'efficacia e sull'efficienza della formazione. L'apprendimento immersivo e personalizzato risolverebbe la sfida di adattare i contenuti a un pubblico eterogeneo, garantendo che ogni discente ottenga il massimo dalla propria esperienza formativa (Chen & Wang, 2020; Hwang & Huang, 2018).

In tal senso, il beneficio più significativo sarebbe costituito dalla capacità di adattarsi ai ritmi individuali di apprendimento. A differenza delle metodologie standardizzate, che spesso rallentano i discenti più veloci o rendono frustrati quelli che necessitano di più tempo, i Serious Game potenziati dall'AI permetterebbero a ciascuno di progredire alla propria velocità ottimale (Brusilovsky, 1996), che in ogni caso potrebbe essere adattata nel tempo. Questo approccio non solo migliora l'esperienza dell'utente, ma assicura una completa padronanza dei contenuti prima di passare ad argomenti più complessi, così come in un gioco si passa al "livello successivo".

Questa personalizzazione mirata si traduce in un miglioramento tangibile delle prestazioni: quando i contenuti sono rilevanti e la difficoltà è ottimizzata in tempo reale, il discente sperimenta un maggiore successo, aumentando la sua fiducia, la motivazione e il coinvolgimento. La combinazione di immersione (data dal Serious Game) e precisione (data dall'AI) favorirebbe una capitalizzazione delle conoscenze superiore rispetto ai metodi tradizionali, poiché l'apprendimento risulterebbe contestuale e pragmatico (Anderson & Schunn, 2000).

Infine, l'approccio integrato porta a una maggiore efficienza della formazione. Che si tratti di un contesto aziendale o accademico, l'AI riduce il tempo necessario per la formazione, identificando rapidamente i punti deboli e concentrando lo sforzo del partecipante ove maggiormente necessario. L'investimento in queste tecnologie si tradurrebbe in un ritorno sia in termini di acquisizione di competenze di alto livello, sia in un'ottimizzazione complessiva dei costi e delle tempistiche della formazione.

Nonostante il potenziale rivoluzionario di cui si è discusso, l'integrazione di Serious Game e AI non è certo priva di sfide (UNESCO, 2019; European Commission, Joint Research Centre, 2021). L'implementazione di questi sistemi richiederebbe grandi investimenti in termini di sviluppo e risorse, sia per la creazione di Serious Game complessi che per l'addestramento dei modelli di Intelligenza Artificiale. La qualità dei dati rappresenta l'elemento cruciale (Siemens, 2013; Baker & Inventado, 2014), dal momento che l'AI è efficace sulla base dei dati su cui è stata addestrata, e ottenere dati di apprendimento pertinenti e sufficienti può essere complesso. Inoltre, la piena integrazione di AI in piattaforme di gioco esistenti o nuove, richiede competenze tecniche avanzate e può presentare ostacoli di compatibilità.

Parallelamente alle sfide tecnologiche, sono presenti sfide rilevanti anche dal punto di vista etico (Punie & Cabra, 2021). L'AI, analizzando dettagliatamente il comportamento degli utenti all'interno dei Serious Game, raccoglie una vasta quantità di dati sensibili. La privacy dei dati dei discenti diventa una questione centrale, richiedendo protocolli di sicurezza rigorosi e politiche di trasparenza chiare sull'utilizzo delle informazioni. È inoltre necessario che i processi decisionali degli algoritmi siano comprensibili e non introducano bias nell'apprendimento per un sistema equo.

Queste sono solo alcune delle sfide che ci potrebbero essere, ma nonostante ciò, le prospettive future restano estremamente promettenti e positive. Inoltre, i progressi nel campo del Natural Language Processing (NLP) e dell'analisi del sentimento permetteranno interazioni ancora più realistiche e sfumate all'interno dei Serious Game. L'AI continuerà a evolvere, offrendo simulazioni sempre più complesse e percorsi di apprendimento iper-personalizzati. Questa convergenza è destinata non solo a migliorare la formazione in contesti specifici, ma a democratizzare anche l'accesso a un'educazione di alta qualità a livello globale (WEF, 2020; OECD, 2021).

## **Conclusioni**

L'integrazione di Serious Game e Intelligenza Artificiale segna un punto di svolta fondamentale nel settore della formazione digitale. Come emerso da questo articolo, tale convergenza supera i limiti della didattica tradizionale, fornendo un approccio che è sia profondamente immersivo che scientificamente personalizzato. I Serious Game offrono il coinvolgimento e l'esperienza pratica, mentre l'AI fornisce l'intelligenza adattiva necessaria per modulare il percorso formativo in tempo reale, in base alle esigenze e ai ritmi di ogni singolo discente. Nonostante le sfide legate all'implementazione e alle questioni etiche, i benefici derivanti da questa sinergia sono innegabili, garantendo maggiore efficacia nell'apprendimento e un'ottimizzazione delle risorse. L'apprendimento personalizzato e immersivo non è più solo una visione teorica, ma una realtà che sta ridefinendo il futuro dell'educazione nell'era digitale. La combinazione di Serious Game e AI è dunque destinata a diventare un pilastro fondante della formazione di nuova generazione, rendendo l'apprendimento più efficiente, pertinente e coinvolgente che mai.

## **Referenze:**

- Anderson, J. R., & Schunn, C. D. (2000). The ACT-R Theory of Cognitive Skill Acquisition. *Psychological Review*, 107(1), 3-33.
- Baker, R. S., & Inventado, P. S. (2014). Educational Data Mining and Learning Analytics. In *Handbook of Learning Analytics and Educational Data Mining* (pp. 61-75).
- Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 6(2-3), 87-129.
- Chen, J., & Wang, Y. (2020). AI-Powered Personalized Learning Systems: A Review. *Journal of Educational Technology Development and Exchange*, 13(1), 1-17.
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining gamification. Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments.
- European Commission, Joint Research Centre (2021). AI in Education: A review of the landscape. Technical Report.
- Gauthier, G., & Frasson, C. (1995). Artificial Intelligence and Education. IOS Press.
- Gee, J. P. (2007). What Video Games Have to Teach Us About Learning and Literacy. Palgrave Macmillan.
- Gomes, F., & Lopes, A. (2021). Adaptive Serious Games: A Literature Review on AI-based approaches. *Journal of Computer Assisted Learning*, 37(4), 1011-1025.
- Hwang, G. J., & Huang, H. C. (2018). AI-driven adaptive learning systems for smart education. *Journal of Educational Technology & Society*, 21(3), 1-14.
- Jonassen, D. H., & Land, S. M. (Eds.). (2012). Theoretical foundations of learning environments. Routledge.

- Ke, F. (2009). A Qualitative Meta-Analysis of Computer Games as a Learning Tool. *Educational Technology Research and Development*, 57(5), 603-625.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.
- Luckin, R., et al. (2016). *Intelligence Unleashed: An argument for AI in Education*. Pearson Education.
- Michael, D., & Chen, S. (2006). *Serious Games: Games That Educate, Train, and Inform*. Thomson Course Technology.
- Nacke, L. (2013). Adaptive Serious Games: The Role of AI in Personalizing Learning Experiences. *AI & Society*, 28(2), 221-230.
- OECD (Organizzazione per la Cooperazione e lo Sviluppo Economico) (2021). *The Impact of Digitalisation on Learning*. Rapporto di ricerca.
- Prensky, M. (2001). *Digital Game-Based Learning*. McGraw-Hill.
- Punie, Y., & Cabra, V. (2021). Artificial Intelligence and the future of Education. European Training Foundation (ETF) Working Paper.
- Ritterfeld, U., Cody, M., & Vorderer, P. (Eds.). (2009). *Serious Games: Mechanisms and Effects*. Routledge.
- Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*, 57(10), 1383-1402.
- Squire, K. (2005). Changing the Game: Serious Games for Education. *Innovate* 1(6).
- UNESCO (2019). Artificial Intelligence in Education: Challenges and opportunities for sustainable development. Rapporto di sintesi.
- WEF (World Economic Forum) (2020). *The Future of Jobs Report*. Rapporto sul futuro delle competenze e della formazione.
- Woolf, B. P. (2009). *Building Intelligent Interactive Tutors: Student-Centered and Scalable Tutoring with AI Technologies*. Morgan Kaufmann.

## **Cyberstalking, perché si odia online: dinamiche e ruolo delle piattaforme – di Iva Marino, PhD in Health Promotion and Cognitive Sciences e Francesca Rizzuto, professoressa associata nel Dipartimento Culture e Società dell’Università di Palermo**

*Una riflessione sulle dinamiche tra logiche delle piattaforme e i circuiti planetari dell’odio, evidenziando alcune significative conseguenze a livello individuale e sociale. In particolare, l’attenzione verrà focalizzata sulla problematica interazione tra esigenze di protezione dei soggetti vulnerabili e le possibilità di attacchi e azioni persecutorie, favorite dalla logica delle piattaforme a livello planetario*

### **I social media come arene per hate speech e stalking virtuale. Alcune conseguenze sociali (di Francesca Rizzuto)**

La rilevanza dei media nelle pratiche di costruzione di senso attivate individualmente e collettivamente è da decenni uno dei temi centrali delle scienze della comunicazione, che ne hanno proposto letture e analisi differenti, pur riconoscendone la sempre maggiore significatività, soprattutto nell’ambito delle dinamiche contemporanee di circolazione planetaria di messaggi. Nella società delle piattaforme (Van Dijck, Poell, de Waal, 2018), le tecnologie della comunicazione rendono possibile la creazione di un sovraccarico simbolico inedito, che si diffonde nell’agorà digitale grazie ai social media, diventati un serbatoio vasto e articolato di risorse cognitive e ambiti di relazioni rilevanti per i singoli. Con l’esperienza quotidiana della pluralizzazione dei mondi e del transito generalizzato in ambienti reali e virtuali (Sorice, 2020; Cappello-Rizzuto 2024), i contenuti mediatici forniscono frame, delimitano orizzonti di significati condivisi da alcuni soggetti che, chiusi in circuiti caratterizzati dalla tendenza all’omofilia e all’eliminazione della dissonanza, incoraggiano l’esplicitazione di pregiudizi e contribuiscono a confermare stereotipi, usati dagli utenti nel processo di comprensione della realtà. In altri termini, le dinamiche virali, tipiche della circolazione dei messaggi nei social, rispondono al forte bisogno dei singoli di una conoscenza condivisa, arrivando a orientarne l’azione e l’interpretazione della realtà politica e sociale sulla base di tipizzazioni (Nielsen-Ganter, 2022).

Nelle pagine seguenti viene proposta una riflessione sulle dinamiche in atto tra le logiche delle piattaforme e i circuiti planetari dell’odio, sentimento che non si qualifica come effetto collaterale ma come un vero e proprio prodotto sistematico delle architetture di potere algoritmico, che sono oggi in grado di produrre significative conseguenze a livello sociale, politico e individuale. In particolare, l’attenzione viene focalizzata sulla problematica interazione tra le possibilità concrete di produrre e diffondere attacchi verbali o discorsi di odio e le esigenze di protezione dei soggetti, soprattutto i più vulnerabili: la violenza psicologica subdola e pervasiva, che può essere potenzialmente veicolata da qualsiasi utente attraverso le tecnologie digitali, altera non solo i tradizionali circuiti comunicativi delle democrazie rappresentative, in direzione di una sempre più marcata tendenza alla polarizzazione ma, a livello individuale, condiziona negativamente la percezione della realtà della vittima-bersaglio di attacchi personali, compromettendone profondamente il senso di sicurezza e le relazioni sociali.

Nell’ecosistema mediale ibrido, uno dei rischi più significativi è quello di accettazione acritica da parte degli individui di una rappresentazione della realtà basata su distorsioni ideologiche e pratiche di disinformazione che polarizzano il dibattito pubblico: pertanto, è utile riflettere sulle dinamiche della relazione causale tra i contenuti veicolati nelle piattaforme e lo sdoganamento di

comportamenti violenti o incivili, non solo nella sfera pubblica ma anche nelle interazioni private che. Molti studi hanno evidenziato il processo di affermazione di varie forme di violenza verbale in differenti contesti nazionali: già nel 2021, Bentivegna e Boccia Artieri, rilevavano che “le pratiche comunicative abilitate dalle piattaforme online, congiuntamente ad una contrapposizione militante e diffusa nei confronti delle élite, rappresentano il carburante che alimenta la diffusione dei discorsi aggressivi da parte dei cittadini” (Bentivegna, Boccia Artieri, 2021, 13). In un contesto sociale che convive con le dimensioni dell’incertezza e dell’insicurezza, tragicamente acute dai numerosi conflitti in corso e dallo scardinamento degli equilibri geopolitici del Novecento, l’agorà digitale contemporanea è oggi pervasa dalla proliferazione delle minacce on line e dell’hate speech in un cortocircuito comunicativo, in grado di diffondere e moltiplicare esponenzialmente contenuti fortemente negativi e divisivi, che contribuiscono a minare sempre di più la relazione di fiducia tra cittadini ed istituzioni, anche perché le piattaforme permettono la disintermediazione dei flussi informativi, che agevola pratiche di diffusione di notizie parzialmente o totalmente false in grado di attivare dinamiche di opinione prevalentemente emotive (Corner, 2017).

In altri termini, in una prospettiva meramente deterministica, le peculiarità odierne delle piattaforme sembrano favorire le “profonde trasformazioni del dibattito pubblico, divenuto sempre più spesso occasione di scontro e posizionamento piuttosto che confronto e discussione” (Bentivegna- Boccia Artieri, op.cit., 7): tuttavia, come premessa indispensabile, ci sembra opportuno evitare l’errore di adottare un approccio interpretativo caratterizzato da un banalizzante riduzionismo comunicativo (Barisione 2020), che deresponsabilizza sia gli attori politici che le imprese mediatiche e spinge a considerare lo sviluppo planetario dei social media come l’unica causa della crisi profonda dei sistemi democratici occidentali, mentre è evidente che vi sono numerosi fattori extra-mediali nella scena globale.

Nelle democrazie occidentali fenomeni come l’aumento dell’astensionismo o l’ascesa di varie forme di populismo devono essere letti e spiegati anche in relazione a tensioni sociali e politiche risalenti ad alcuni decenni fa, vale a dire a contesti comunicativi pre-social; indubbiamente, però, nell’era digitale ci sono elementi inediti rispetto al passato, che hanno modificato i palcoscenici comunicativi e le interazioni sociali (Chadwick, 2013), frammentando i flussi informativi e moltiplicando gli attori in grado di produrre e diffondere news. Agendo accanto al prevalere di logiche di mercato, tutti questi fattori hanno rafforzato le tendenze alla rappresentazione polarizzata della realtà, anche quella politica, e accelerato le pratiche narrative fondate sull’uso della componente emotiva come lente per “raccontare” la realtà e selezionarne eventi e attori da raccontare (Papacharissi, 2015). La peculiarità più preoccupante dell’ecosistema digitale contemporaneo è quella di diventare troppo spesso una potente “fabbrica dell’odio”, in grado di produrre potenti effetti di realtà (McIntyre, 2018) attraverso pratiche di condivisione di notizie divisive, caratterizzate dal ricorso a scelte lessicali improntate all’odio e alla violenza verbale.

L’esito più preoccupante è la trasformazione tossica del dibattito democratico, realizzata attraverso un continuo appello alla sfera emotiva dei singoli, al fine di suscitare rabbia e odio da canalizzare in precise strategie politiche. Distaccandosi dalla razionalità e dalle pratiche argomentative fondate sul dialogo, si è diffusa una cultura dell’odio che, grazie ai processi di viralizzazione, può annientare individui, leader, perfino intere popolazioni, scardinando le basi stesse del diritto alla protezione della privacy, tradizionalmente riconosciuto come fondamentale in tutti gli ordinamenti democratici (Rizzuto, Sciarrino 2021). Come già evidenziato da numerosi autori Balzerani (2019), la dimensione emotiva influenza profondamente il vivere sociale in quanto le emozioni sono leve basilari del

comportamento individuale e di gruppo, che influenzano sfere apparentemente lontane tra loro come la politica, l'informazione, le relazioni tra etnie o religioni: il problema contemporaneo è che le emozioni, nel contesto digitale, sono usate come strumenti di costruzione dei messaggi a livello politico o per attrarre destinatari delle news, perché in grado di intercettare e rafforzare facilmente audience e consensi, indispensabili nella logica commerciale dominante nel mondo mediale e nel contesto della politica-spettacolo.

La centralità della logica emozionale è rinvenibile anche nell'informazione che, da almeno tre decenni, ha privilegiato la prospettiva spettacolare in un contesto rivoluzionato sul piano tecnologico e stravolto su quello economico dall'affermazione di nuovi attori che producono e diffondono news: con l'emotainment (Santos 2009), già prima dell'avvento dei social, la narrazione giornalistica della realtà si era focalizzata sulla capacità di emozionare i riceventi per catturarne l'attenzione, attraverso il ricorso a criteri di notiziabilità attenti alla componente drammatica e conflittuale degli eventi. Tuttavia, la teatralizzazione della realtà narrata è diventata il tratto peculiare delle pratiche informative contemporanee, che hanno trovato, nel contesto delle piattaforme, nuova linfa capace di alimentare e rafforzare i processi di patemizzazione e vetrinizzazione del sociale, dando vita a racconti-resoconti verosimili, opachi, banalizzanti in cui la verità oggettiva dei fatti diventa, coerentemente con la logica della post-verità (Joux, 2023; Cappello-Rizzato 2024), sempre meno rilevante rispetto alla capacità di suscitare emozioni forti (Marinov 2019).

Sul versante politico, l'appello alle emozioni a cui i leader hanno sempre fatto ricorso, pur con declinazioni differenti, è oggi la leva comunicativa più usata in contesti politici anche molto lontani, con l'obiettivo di suscitare consensi presso i cittadini-elettori sia durante le fasi elettorali che nel corso delle attività routinarie di governo. Se nella società contemporanea ogni luogo è diventato palcoscenico, grazie ai social e alla pervasività degli schermi, la dimensione dell'intrattenimento è ormai componente fondamentale della cultura dell'esposizione, ma questo continuo appello alle emozioni presenta alcuni rischi: infatti, un circuito comunicativo prevalentemente emozionale permette di attrarre, più facilmente e senza intermediazioni, attenzione e consensi nella sfera pubblica piattaformizzata (Sorice, 2023), ma preferire far leva sulla componente emotiva, a scapito di modalità informativo-cognitive che permettono di comprendere temi o priorità, rende alcuni attori istituzionali e professionisti dell'informazione corresponsabili non solo nei processi di diffusione dell'odio e della violenza verbale ma anche della normalizzazione di tali pratiche discorsive divisive nelle interazioni quotidiane. La pervasività di modalità comunicative conflittuali e dell'hate speech può essere letta, in tale prospettiva, come l'esito planetario più disastroso di un'attività cumulativa e pianificata di "nutrimento dell'odio nei riguardi dei Nemici", realizzata e propagandata soprattutto attraverso narrazioni diffuse nei social. Nel mondo digitale le storie proposte nella rete, fatte circolare anche dai media mainstream, offrono sempre più spesso visioni del mondo e modelli di odio facili da comunicare, che attivano un pericoloso contagio emotivo e forniscono ai leader una piattaforma semplificante e facilmente accessibile per dialogare "direttamente" con gli individui.

L'inciviltà contemporanea del dibattito pubblico (Boccia, Bentivegna 2021), dagli interventi di Trump o di Putin fino ai dibattiti sull'antisemitismo, conferma che l'ecosistema mediale ibrido svolge senza dubbio un ruolo chiave per la diffusione di contenuti violenti e di odio, perché nelle piattaforme prevalgono pratiche comunicative caratterizzate da brevità, sinteticità, linguaggio aggressivo e volgare, che ricorrono ad una logica rigidamente binaria (noi/loro): facendo leva sul senso di appartenenza le fonti di messaggi di odio sono favorite dalla possibilità concreta di sfuggire a eventuali fonti di

dissonanza cognitive garantita dalle relazioni nelle echo chambers o nei gruppi on line (Riva, 2018) e dall'inconsistenza di sanzioni normative efficaci che caratterizza i sistemi legislativi di molti paesi.

Una delle conseguenze più significative è una problematica ed evidente alterazione della dinamica della discussione pubblica, nella quale viene data forza ad alcuni o, al contrario, si attivano meccanismi di spirale del silenzio per altri individui o gruppi: attraverso i media si può monitorare il contesto e si favorisce una crescente tensione tra posizioni maggioritarie o minoritarie (o ritenute tali), che spingono all'ostilità verso l'altro, erodendo il valore del confronto aperto tra posizioni differenti, con enormi ricadute sul comportamento elettorale e sulla fiducia dei singoli nei confronti della politica. Il rischio maggiore è che “possa prevalere una cultura del disprezzo capace di aumentare l'odio verso Altri-Nemici” (Balzerani, 2019) e favorire quel processo di escalation emotiva dalla rabbia al disgusto, che nella storia recente dell'Europa, ha portato a catastrofi immani. Ecco perché la diffusione odierna delle varie forme di inciviltà deve preoccuparci come studiosi e come cittadini: non solo l'uso di linguaggi aggressivi e offensivi da parte dei leader, ma anche la normalizzazione di vere e proprie strategie retoriche dell'insulto, usato come arma politica per screditare l'avversario, possono inesorabilmente condurre all'accettazione di un clima generale di intimidazione verso chi dissente, in cui l'Altro viene descritto con il linguaggio dell'odio, ricorrendo a scelte lessicali deformanti, stili espositivi ipersemplificati e narrazioni distorte.

### **La logica delle piattaforme e gli effetti psicologici individuali (di Iva Marino)**

Viviamo un'epoca a conduzione narcisistica, caratterizzata da fluidità, contraddizioni, evidente è una tensione interiore tra un estremo bisogno dell'altro e la realizzazione di un implacabile individualismo. In questo scenario è diffusa una cornice di incertezza e frammentazione, connessa ad un “relativismo valoriale” che sembra aver generato un sempre più frequente ripiegamento narcisistico sul sé. Il mondo moderno, nel quale i bisogni individuali sono considerati di primaria importanza, incoraggia le persone a concentrarsi su di sé, inducendo a un torpore che talvolta genera carnefici e mostruosità. La nostra società, come la nostra vita, spinta da due forze specifiche (Eros, Dio greco dell'amore che rappresenta l'istinto di vita o la pulsione di conservazione, l'energia, la connessione, Thanatos, Dio greco della morte che rappresenta la pulsione di distruzione), è fortemente Thanatos e poco Eros, rivolta così alla violenza e all'autodistruzione; si avverte, inoltre, una ridotta sensibilità, un impoverimento culturale, che induce precarietà come condizione esistenziale e psicologica. In questo quadro alquanto controverso, in cui i social media hanno trasformato le modalità di comunicazione e creato nuove forme di interazione sociale, la velocità di diffusione dei contenuti, imponendo la logica dell'engagement algoritmico che, grazie all'anonimato, crea un ambiente fertile per la diffusione non solo dell'odio verso gruppi o etnie, ma anche delle molestie rivolte a bersagli specifici. Nel tessuto sociale contemporaneo, la soggettività di ciascuno di noi, ciò che Wallace (2001) definisce “persona on line”, vive immersa in un continuum di esperienze digitali e reali, che influenzano ogni aspetto dell'identità individuale, come la percezione del Sé, la relazione fra il Sé digitale e il Sé reale, l'affettività così come la sessualità e la socialità, fino ai casi patologicamente rilevanti di vera e propria dipendenza dalle tecnologie o di abusi su altri individui. Del resto, l'ambiente digitale è tutt'altro che neutrale: la sua architettura, orientata a incentivare l'interazione costante, muove, spesso, contro il benessere individuale e contro la coesione sociale: tra le espressioni più gravi e meno visibili vi è lo

stalking virtuale, una forma cieca e crudele di violenza psicologica, che colpisce donne, minoranze ed anche figure pubbliche.

All'interno dei social, la violenza è un diventata modello di comportamento in una società “archetipica dell'aggressività”, nella quale si avverte una frequente scissione fra il “pensare ed il fare”, una mancata risonanza emotiva che diviene veicolo di odio. Nel contesto delle piattaforme due fattori favoriscono la circolazione di odio e violenza: la presenza delle echo chambers, spazi informativi chiusi in cui le opinioni si rafforzano per effetto della omogeneità cognitiva (Sunstein, 2001) e l'uso delle tecnologie per monitorare, intimidire o controllare un individuo. Sebbene apparentemente distinti, entrambi i fenomeni condividono una base comune: l'infrastruttura algoritmica e relazionale delle piattaforme digitali. In queste pagine viene proposta una riflessione critica ed interdisciplinare sull'interazione fra le logiche algoritmiche delle piattaforme e la proliferazione dell'odio on line, con particolare attenzione al cyberstalking, analizzandone le conseguenze più significative a livello individuale, evidenziando un modello del trauma, i rischi per la salute mentale, tra esigenze di protezione dei soggetti vulnerabili e le possibilità di attacchi ed azioni persecutorie favorite dai media digitali, mettendo in luce implicanze sociologiche, psicologiche. Nelle echo chambers gli utenti sono esposti solo a opinioni compatibili alle proprie: viene, quindi, evidenziata una visione univoca su un argomento, in cui le informazioni, le idee o convinzioni dei singoli vengono rafforzate dalla ripetizione all'interno di un sistema definito (Sunstein, 2001) e le visioni diverse o concorrenti sono censurate o non accolte: ciò determina la creazione di bias cognitivi e la diffusione di tecniche di proliferazione e di marketing, in quanto, una volta che si acquisiscono le informazioni relative ad un utente, ad esempio di ciò che lui sceglie o preferisce, è semplice inviare messaggi persuasivi e cancellare altri interessi.

Da quel momento, gli algoritmi dei social network tenderanno a mostrare messaggi, commenti e notizie verso i quali, in precedenza, un individuo ha già mostrato interesse. Il meccanismo delle camere d'eco alimenta una progressiva radicalizzazione ideologica, in cui l'odio non è più percepito come una semplice deviazione, ma come parte integrante dall'identità collettiva di un gruppo. In questo contesto, il discorso d'odio si trasforma in un contenuto altamente performante, poiché amplificato autonomamente dalle dinamiche algoritmiche che privilegiano l'engagement. Un ambiente digitale così strutturato favorisce la formazione di comunità virtuali “chiuse, polarizzate”, dove questo sentimento non rappresenta un effetto collaterale, ma un prodotto sistematico delle architetture del potere algoritmico. Le camere d'eco implementano la progressiva normalizzazione di linguaggi aggressivi, dove “l'odio”, per sua natura emozione forte e potente, diviene uno dei prodotti più redditizi. La selezione algoritmica e la tendenza all'omofilia sociale rendono queste camere dell'eco un fenomeno strutturale, che riduce l'apertura al dissenso, inibisce la riflessione critica e accentua la polarizzazione affettiva: in molti casi l'ostilità verso chi ha opinioni diverse non nasce solo da divergenze ideologiche ma da reazioni emotive di repulsione e l'esposizione ripetuta non solo rafforza convinzioni rigide ma indebolisce progressivamente la tolleranza sociale.

La permanenza in echo chambers può generare, quindi, rigidità cognitiva, intolleranza all'ambiguità, una sovraesposizione emotiva ai contenuti confermativi: in generale, le camere dell'eco riducono la complessità del pensiero e generano meccanismi di gruppo che rinforzano l'ansia identitaria ed il bisogno di appartenenza. La logica delle piattaforme digitali non è mai neutra: come già evidenziato, i social media operando all'interno di una “economia dell'attenzione”, in cui ogni interazione digitale viene trasformata in valore economico, diffondono anche i contenuti d'odio, purché non violino formalmente le policy, proprio perché possono essere economicamente vantaggiosi (Citron, 2014): la

conseguenza più problematica è una evidente lentezza strutturale delle piattaforme nel rimuovere contenuti violenti o discriminatori, specialmente quando questi generano traffico, vale a dire quando il perseguitamento del profitto prevale sulla responsabilità etica (Bush, 2023). Non manifestando alcun interesse per un problema comportamentale, ma supportando esclusivamente “una visione” di scelte tecnologiche e di “dinamiche” di engagement (Gillespie, 2018) che governano gli algoritmi di raccomandazione, i social si basano, del resto, su un modello di business caratterizzato dalla difficoltà di applicare una moderazione efficace, soprattutto quando questa si scontra con ostacoli tecnici, linguistici, culturali: in altri termini, le intelligenze artificiali, nel filtrare i contenuti, sono spesso incapaci di comprendere il contesto o l’uso strategico di codici linguistici criptici e, a causa di questo, concorrono alla diffusione sistematica dell’odio. Il legame fra architetture digitali e violenza simbolica è una questione centrale nel dibattito politico e sociale dell’ultimo decennio: numerose piattaforme digitali come Facebook, Tik Tok, Instagram e YouTube utilizzano “algoritmi predittivi” per proporre contenuti che massimizzino il coinvolgimento emotivo dell’utente e tali algoritmi, servendosi di metriche relative a like, condivisioni e tempo di permanenza, finiscono spesso per privilegiare contenuti estremi, sensazionalisti poiché questi attirano maggiormente attenzione (Tufekci, 2015). In altri casi, si realizza un effetto finestra sul cortile, quando molte piattaforme digitali diventano “finestre virtuali”, attraverso cui spiare o osservare vite altrui così come eventi violenti o drammatici, spesso condivisi in tempo reale, fra un mix di orrore ed attrazione.

#### **Lo stalking virtuale come violenza digitale e psicologica: cause e ricadute sulla vittima** (di Iva Marino)

È ancora controversa la questione relativa al fatto che il cyberstalking rappresenta un’entità autonoma (McFarlane et al., 2003) o piuttosto un’estensione dello stalking tradizionale, per cui la molestia online è generalmente associata a quella off line (Burgess & Baker, 2002). In base al grado di sovrapposizione tra stalking e cyberstalking, Spence-Diehl (2003) distinguono tre categorie: cyberstalking limitato al cyberspazio, cyberstalking che inizia nel cyberspazio per estendersi successivamente offline e cyberstalking associato ad altre modalità di stalking tradizionale. È utile distinguere, a proposito della specificità del cyberstalking, l’online harassment o “molestia occasionale” dal cyberstalking vero e proprio che è delineato da azioni sistematiche di minacce e molestie che si protraggono nel tempo. Harvey (2021), relativamente a ciò, parla di cyberstalking diretto, caratterizzato dall’impiego degli strumenti di comunicazione elettronica come una modalità per stabilire e mantenere il contatto con la vittima, e cyberstalking indiretto nei casi in cui l’uso degli strumenti di comunicazione mira a diffondere informazioni false o private, sottoscrivere a nome della vittima servizi informatici, assumere la sua identità in discussioni private o pubbliche. Mentre il cyberstalking diretto si mantiene nei limiti di una dimensione privata della relazione con minacce, molestie rivolte direttamente alla vittima o danni alle sue applicazioni o dispositivi informatici, quello indiretto coinvolge sempre una terza parte, assume una dimensione pubblica, coinvolgendo famiglia, amici, colleghi e datori di lavoro della vittima, con l’intento dello stalker di danneggiarne la reputazione. Con la diffusione rapida di nuovi sistemi di comunicazione e di partecipazione online, è aumentata in maniera significativa la possibilità di perpetuare minacce, rischi di manipolazione dell’identità e delle relazioni, un’enorme quantità di dati disponibili che il cyberstalker non ha difficoltà a reperire. Non ha difficoltà a conoscere e riconoscere abitudini, relazioni della “potenziale vittima”.

L’utilizzo dei nickname, indirizzi e-mail multipli e sistemi di re-mailer proteggono l’identità del cyberstalker, riducendo il controllo sociale e favoriscono l’espressione di comportamenti antisociali. Uno studio condotto su 24 soggetti, rifacendosi alla classificazione delle condotte di stalking proposta

da Mullen et al. (2000), Mcfarlane e Boccia (2003) distinguono le seguenti tipologie: 1) vindictive cyberstalker, motivati dal desiderio di vendetta, con una cultura informatica medio alta, con una conoscenza ad ampio spettro di tecnologie informatiche, compresi lo studio degli algoritmi, di sistemi di monitoraggio a distanza, di intrusione nel sistema operativo della vittima; 2) aggressivi, una tipologia spesso con problemi e precedenti psichiatrici e penali; 3) i composed cyberstalker, soggetti che non presentano generalmente precedenti psichiatrici né penali, ma, semplicemente, sono spinti dal piacere di infastidire e recare molestia alle loro vittime, con cui non hanno stabilito precedentemente né cercano di stabilire un contatto offline. Infine, 4) gli intimate cyberstalker, ovvero coloro che ricercano attenzione e gratificazione sessuale (in generale sono gli ex partner, che rivolgono alle vittime tentativi disperati di riconciliazione, oppure rivolgono un'intrusiva attenzione a persone appena conosciute, di cui soffrono di una infatuazione estemporanea) e 5) i collective cyberstalker, che comprende tutti i casi di molestie perpetrati da un gruppo di soggetti e non da un singolo stalker. Il fenomeno descritto è in continua espansione ma quale è il profilo dello stalker? Quali gli aspetti comportamentali specifici?

Lo stalker manifesta sempre, in generale, un'evidente problematica nell'aerea affettivo- emotiva, relazionale e comunicativa a cui non sempre corrisponde un preciso quadro psicopatologico. Interessante è delineare gli stili di attaccamento nei confronti delle vittime: parliamo di "attaccamento affettivo-amoroso ed attaccamento persecutorio- irato"; un tipico attaccamento insicuro- ambivalente- ansioso (J. Bowlby, 1969), è tipico dei comportamenti di stalking. Gli individui che presentano questi pattern si caratterizzano per l'ansia nelle relazioni mettendo in atto comportamenti associati a gelosia, rabbia e rivendicazione verso il partner; comportamenti intrusivi, molesti e persecutori. Mullen e Purcell (2000) hanno creato un campione significativo di valutazioni cliniche in cui si riscontra, secondo un approccio multidimensionale, una relazione significativa fra comportamenti di stalking e disturbi della personalità. Nella classificazione effettuata, il "risentito" rappresenta, di solito, un ex-partner che desidera vendicarsi per la rottura della relazione sentimentale causata, a suo avviso, da motivi ingiusti: agisce ledendo direttamente la persona, la sua immagine o le sue proprietà (casa, macchina, ecc). È il tipo di stalker che "pubblica sul web foto o immagini private" (cyberstalking), aspetta fuori casa o segue la vittima, danneggia la sua macchina, ecc: questo è il tipico esempio di soggetto con disturbo narcisistico di personalità, che muove nella specificità fra il "bisognoso d'affetto" ed il molestatore assillante, dove ogni segnale di vicinanza o di confidenzialità espressa dalla vittima viene interpretato come una naturale espressione del desiderio di contatto. Specificatamente, un narcisista vulnerabile, di tipo "covert" che si manifesta nella figura del "corteggiatore impacciato", inadeguato ad entrare in relazione con la vittima, che esperirebbe grandemente oppressione e violenza psicologica on line, per destare attenzione e controllo sul bersaglio.

Ma quali le cause dello stalking? Oltre agli stili di attaccamento, come si è detto, determinante è la disregolazione affettiva, la mancata capacità di mentalizzare, l'ansia da separazione, la vergogna, condizionata dalla svalutazione della vittima e, soprattutto, la "dipendenza relazionale". E se "il legame desiderato" si trasforma in un'ossessione che domina la mente, una questione pertinente alle emozioni, a pensieri intrusivi, allora si manifesta una ricerca compulsiva dell'oggetto d'amore a cui si lega una vasta gamma di elementi disfunzionali, una condizione senza la quale l'esistenza appare perdere significato.

Un vissuto emotivo caratterizzato dall' incertezza, dal disorientamento, dalla fragilità e dall' angoscia di far svanire: uno stato patologico allorché la relazione è vissuta come condizione unica, indispensabile, connessa ad un giusto equilibrio fra vicinanza e separazione poco assimilato nel tempo magari e ad una mancata la presenza di un genitore supportivo, rassicurante, sensibile. La vittima perseguitata manifesta sensazioni ed emozioni intense, che vanno da uno stato di allerta e di stress psicologico a stati di preoccupazione, di paura, di disprezzo per il molestatore. Violata la dimensione personale e privata, intenso è il disturbo post traumatico da stress, le vittime esperiscono un tormento devastante e subiscono conseguenze anche molto gravi: nel caso dello stalking virtuale, infatti, i danni psicologici possono essere comparabili, oltre a quelli del disturbo post traumatico da stress PTSD, allo sviluppo di sintomi di iper-vigilanza, paranoia, isolamento, angoscia e calo dell'autostima (DSM 5th ed., 2013). Più a rischio sono alcuni gruppi vulnerabili: ad esempio, i minori per la loro scarsa consapevolezza, le persone LGBTQIA+ per il loro orientamento sessuale, a rischio di outing forzato, le donne giovani fino a 35 anni con scarso senso di agency, coping, resilienza che sviluppano maggiormente un PTSD; ovviamente, tra le vittime più frequenti vi sono anche personaggi pubblici o influencer per frame relativi all'invidia narcisistica o all'osessione o delirio di relazione del persecutore. Per descrivere la narrativa della vittima che subisce molestie dobbiamo introdurre il concetto di condizione traumatica, ovvero un significativo stato di compromissione della salute mentale. Numerose le definizioni, nella storia della psicopatologia, del termine trauma che rappresenta un evento spartiacque, una frattura emotiva che mina il senso di stabilità, di sicurezza e di identità, di continuità fisica e psichica.

Sebbene il senso di sicurezza sia a molti cosa nota, risulta difficile definirlo ma scientificamente bisogna focalizzarsi soprattutto sul concetto di accoglienza. L'accoglienza è il contenuto portante per giungere alla definizione di sicurezza. Il sentirsi "accolti" specificamente "riconosciuti" procura nel tempo uno stato affettivo di benessere; all'opposto uno stato di rischio e di paura perdurante, per una costellazione di comportamenti intrusivi, assillanti, di azioni che si ripetono nel tempo, con frequenza ed intensità diverse, rappresenta una forma di violenza psicologica subdola e pervasiva che sfrutta la tecnologia per alterare la percezione della realtà della vittima, compromettendone profondamente il senso di sicurezza personale. Questo tipo di trauma può lasciare segni profondi, non solo a livello cognitivo (memorie traumatiche) ma anche affettivo, neurobiologico e corporeo, una dimensione del trauma che invade lo spazio del tempo, della coscienza, del pensiero e delle emozioni (Sapiro, 2016).

Comprendere il PTSD, in relazione al cyberstalking, richiede dunque un approccio approfondito che consideri il funzionamento delle memorie traumatiche compromesse (Arcipelago M- Molestia), la mente affettiva descritta da Panksepp (1998) ed i principi della teoria polivagale di Porges (2011). Un passato che non passa: le vittime di cyberstalking spesso riportano esperienze intrusive e ricorrenti che non si limitano al periodo della persecuzione, ma continuano a vivere nella mente e nel corpo. Le memorie traumatiche non vengono integrate correttamente nella narrazione autobiografica della persona, invece di essere archiviate come ricordi del passato rimangono attive, frammentate, dissociate. Questo compromette la capacità dell'individuo di distinguere tra "allora" e "adesso", innescando risposte di allerta cronica, ipervigilanza, flashback anche in contesti sicuri. Nel cyberstalking la natura persistente, spesso anonima, dell'aggressione amplifica il senso di impotenza: il cervello registra questa esperienza come minacce esistenziali e le memorie associate non seguono i normali processi di consolidamento, rendendo difficile l'elaborazione ed il superamento del trauma.

Jaak Panksepp, psicologo e neuroscienziato affettivo, che ha descritto il cervello come strutturato in sistemi emotivi primari, presenti anche negli animali, che mediano la nostra esperienza affettiva di

base, offre un valido contributo per comprendere la natura dei sintomi del PTSD. Interessante sottolineare che questi sistemi di cui parla includono, tra gli altri, la paura, la ricerca, il panico/separazione e la rabbia. Nelle vittime di cyberstalking il sistema della paura è cronicamente iperattivato: ogni notifica, messaggio o segnale digitale può riattivare una risposta emotiva sproporzionata perché la mente affettiva registra questi segnali quale “pericolo”, anche in assenza di una minaccia oggettiva. Il sistema panico/separazione, legato appunto alla separazione e all’abbandono, potrebbe anche essere coinvolto, specie se lo stalker è una persona con la quale si aveva un legame. Le esperienze vissute non sono solo cognitive ma vissute come “visceralmente emotive” e si imprimono profondamente nella psiche ed il trauma colpisce il cervello emotivo. Stephen Porges, con i suoi studi, ha trasformato la nostra comprensione della risposta “allo stress e al trauma”: secondo questo autore, il sistema nervoso autonomo si regola su tre modalità principali: il sistema ventro-vagale, che esprime connessione, sicurezza, attivo in condizioni di calma e relazione; il sistema simpatico che esprime le azioni di attacco o fuga e si attiva in condizioni di pericolo ed il sistema dorso-vagale, che esprime collasso o congelamento ed entra in funzione quando la minaccia è percepita come ineluttabile. Nel PTSD la persona può restare bloccata in uno stato di iperattivazione (sistema simpatico) o di spegnimento (sistema dorso-vagale): questo spiega i sintomi delle vittime di cyberstalking apparentemente opposti come l’ansia intensa e la dissociazione, l’irritabilità e la sensazione di intorpidimento emotivo. La teoria polivagale sottolinea il ruolo cruciale del “nervo vago” nella regolazione emozionale, nella sensazione di sicurezza.

Quando questa regolazione viene meno a causa di un trauma, l’individuo può sentirsi perennemente in uno stato di allarme o di isolamento interiore. Il PTSD non è, quindi, soltanto un insieme di pensieri negativi ma produce sintomi reali, invisibili e dolorosi come la difficoltà nella regolazione emotiva, alcune sensazioni corporee anomale (tensioni, vuoti, blocchi) o una sensazione cronica di insicurezza anche in ambienti sicuri, fino a conseguenze devastanti sul piano relazionale e sociale come il vero e proprio ritiro sociale, la fatica a costruire legami di fiducia o alcune alterazioni dell’identità e dell’autostima. Tutti questi sintomi sono il risultato di una mente che cerca di “sopravvivere a un’esperienza disturbante”, dove il corpo non mente e l’anima, perseguitata dall’ombra, si ritira e smette di respirare. L’ombra come un vicolo angusto, una porta stretta, diventa dolorosa costrizione che non lascia passare nulla che non sia superfluo (C. Jung, 1951).

### **Odio e responsabilità. Alcune osservazioni conclusive (di Francesca Rizzuto)**

Nell’era delle guerre mondiali a pezzi così come dinanzi alle sfide imposte da alcuni leader ai tradizionali equilibri geopolitici, la cultura dell’odio ha trovato linfa nei circuiti relazionali dei social media, nei quali già la pandemia e, ancora prima, gli attentati terroristici in Occidente avevano reso problematica la definizione stessa e la percezione del sentirsi sicuri. Nella platform society nuovi attori, nuovi gruppi, nuove possibilità tecnologiche hanno contribuito a favorire lo stile comunicativo dell’attacco, vale a dire l’annullamento dell’accezione tradizionale del dialogo come confronto razionale e rispettoso tra opinioni e visioni del mondo. Il panorama contemporaneo dell’ecosistema digitale rende, pertanto, ormai inevitabile la comprensione delle dinamiche comunicative della rete, nella quale il sentimento dell’odio non rappresenta soltanto un effetto collaterale, ma un prodotto sistemico delle architetture di potere algoritmico.

Con l’affermazione di dinamiche comunicative conflittuali che si attivano tra i soggetti, i discorsi d’odio sono diventati un frame comunicativo diffuso sia per i singoli connessi attraverso i social, che per gli attori politici e sociali. Sul piano politico, confinando i cittadini dentro bolle e camere d’eco, in

cui è sempre più difficile entrare in contatto con punti di vista alternativi, i processi di polarizzazione politica e comunicativa sono diventati fattori centrali nei processi di costruzione del consenso, sempre più problematici e volatili nell’ambito dei precari equilibri geopolitici internazionali. Sul piano psicologico individuale, la violenza subdola e pervasiva che sfrutta la tecnologia per alterare la percezione della realtà della vittima, produce danni molto gravi su migliaia di soggetti, compromettendone profondamente il senso di sicurezza personale e la qualità della vita. Da qui l’obbligo per studiosi e attori istituzionali (non solo politici ma anche professionisti dell’informazione) di cogliere, limitare e sanzionare tutti i segnali di odio, non solo per rilevare potenziali pericoli, ma anche per favorire l’instaurarsi di una solida relazione di fiducia nella vita sociale. In un contesto in cui le tecnologie rendono possibile, in una misura planetaria e pervasiva a cui nessuno può sottrarsi, la circolazione (per ragioni diverse e su scala differente) di attacchi a individui, gruppi o etnie per ridicolizzarli o perfino distruggerli, occorre ribadire le responsabilità dei produttori di contenuti on line e definire nuove regolamentazioni condivise a livello internazionale poiché i tradizionali confini dello stato-nazione novecentesco risultano assolutamente inadeguati per elaborare misure di contenimento di fenomeni come l’hate speech o il cyberstalking.

L’articolo è frutto di collaborazione e dialogo tra le due autrici. Ai fini dell’attribuzione, Francesca Rizzato ha scritto i paragrafi 1 e 4; Iva Marino ha scritto i paragrafi 2 e 3.

## Bibliografia

- American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). American Psychiatric Publishing
- Balzerani A., Il codice dell’odio. Segnali non verbali di aggressività e odio nelle relazioni tra gruppi, Roma, Armando ed., 2019.
- Barisione M. (2020), “Contro il comunicazionismo. Per una critica del riduzionismo comunicativo”, in Comunicazione Politica, 3, pp.347-370.
- Bentivegna S., Boccia Artieri G. (2021), Voci della democrazia. Il futuro del dibattito pubblico, Il Mulino, Bologna.
- Bowlby, J. (1969), Attachment and loss. Vol. 1. Attachment, Basic Books, New York.
- Bush, M. (2023), Etica della responsabilità e piattaforme digitali, Laterza, Roma.
- Cappello G., Rizzato F. (2024), Giornalismo, post-verità e media education, Mimesis, Milano.
- Chadwick A. (2013), The hybrid media system. Politics and power, Oxford University Press, Oxford-New York.
- Citron, D. K. (2014), Hate crimes in cyberspace, Harvard University Press, Cambridge (Ma).
- Corner J. (2017), “Fake news, post-truth and media-political change”, Media, Culture & Society, 39 (7), 1100–1107.

- Gillespie, T. (2018), *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*, Yale University Press, New Haven, Connecticut.
- Harvey, D. (2021), *Cyberstalking: Diretto e indiretto*. Franco Angeli, Milano.
- Joux, A. (ed.) (2023). *Journalisme et post-vérité. Les Essentiels d'Hermès*, Paris
- Jung, C. G. (1951), *Aion. Ricerche sul simbolismo del Sé*, Bollati Boringhieri, Milano. (ed. orig. 1951).
- Marinov R. (2019), Mapping the infotainment literature: current trajectories and suggestions for future research, *The Communication Review*, vol, 23, 1-28.
- McFarlane, L., & Boccia, A. (2003), *Studio sulle categorie di stalking, analisi su 24 soggetti*, Franco Angeli, Milano
- McIntyre L. (2018), *Post-truth*, The Mit Press, Cambridge, (Ma).
- Mullen, P. E., Pathé, M., & Purcell, R. (2000), *Stalkers and their victims*, Cambridge University Press, Cambridge.
- Nielsen R. K., Ganter S.A., (2022), *The Power of Platforms: Shaping Media and Society*, Oxford University Press, New York.
- Panksepp, J. (1998), *Affective neuroscience: The Foundations of Human and Animal Emotions*, Oxford University Press, Oxford.
- Papacharissi, Z. (2015). Affective publics and structures of storytelling: sentiment, events and mediality, *Information, Communication & Society*, 19(3), 307-324Riva G. (2018), *Fake news*, il Mulino, Bologna.
- Porges, S. W. (2011). *The Polyvagal Theory: Neurophysiological Foundations of Emotions, Attachment, Communication, Self-Regulation*. W. W. Norton & Company, New York.
- Rizzuto F. (2019), Reality versus Emotions in Italian Journalism, in «*SOFT POWER, Revista Euroamericana de teoria, historia de la politica y del derecho*», vol. 6, pp. 229-245.
- Rizzuto F., Sciarrino V. (2021), “Le fake news ed il diritto all’oblio nell’era della rete. Sulla nuova relazione tra pubblico e privato” in *Sociologia del diritto*, n. 1, pp. 93-118.
- Santos, J. (2009), *Daring to feel. Violence, the Newsmedia and their Emotions*, Lexington Books, Lanham.
- Schudson M. (2003), *The Sociology of News*, W. W. Norton & Company, New York.
- Sorice M. (2020), *Sociologia dei mass media*, Carocci, Roma.
- Sorice M. (2023), Piattaformizzazione della sfera pubblica, *Comunicazione Politica* 3/2020, pp. 371-388, Il Mulino, Bologna.

Spence-Diehl, E. (2003), Cyberstalking and the law: Definition and types, Nova Science Publishers, Hauppauge (NY).

Sunstein, C. R. (2001), Echo chambers: Bush v. Gore, Impeachment, and Beyond. Princeton University Press, Princeton, NJ.

Thussu D.K. (2007), News as Entertainment: The Rise of Global Infotainment, Sage, New York.

Tufekci, Z. (2015) “Algorithmic Harms beyond Facebook and Google: Emergent Challenges of Computational Agency” in Colorado Technology Law Journal, 13(1), 203-218.

Van Dijck J., Poell T., de Waal M. (2018), The Platform Society: Public Values in a Connective World, Oxford University Press, Oxford.

Wallace, P. (2001). The Psychology of the Internet. Cambridge University Press.

## **Dispotismo e terrapiattismo: la deriva culturale del digitale - di Marino D'Amore**, Docente di Sociologia generale presso Università degli Studi Niccolò Cusano

*Il dispostismo digitale trasforma cittadini in consumatori passivi di contenuti filtrati. Algoritmi e piattaforme concentrano potere informativo, sostituendo il pensiero critico con narrazioni semplificate che minacciano le basi della democrazia contemporanea*

L'articolo propone un'analisi di due fenomeni apparentemente distinti, ma strettamente connessi: il *dispostismo digitale*, inteso come delega di funzioni cognitive, decisionali e politiche a sistemi tecnologici, e il *terrapiattismo digitale*, interpretato come la diffusione legittimata di narrazioni semplificate, spesso cospirative, che riducono la complessità del mondo a schemi semplicistici e rassicuranti, affermando come la banalizzazione polarizzata possa favorire la regressione culturale. Il dibattito pubblico sull'innovazione digitale è stato focalizzato per lungo tempo su promesse di emancipazione, democratizzazione dell'informazione e ampliamento delle capacità cognitive collettive, tuttavia, l'esperienza pratica mostra un altro aspetto: la tecnologia da vettore di libertà finisce per accentrare potere verso i medesimi soggetti (big tech, stati, ecosistemi proprietari), mentre gli individui perdono capacità decisionali, abbandonando pratiche critiche in favore di scorciatoie funzionali. Tale delega, voluta, subita o inconsapevole, incarna nel fenomeno del dispostismo digitale. Parallelamente, la cultura digitale veicola forme di semplificazione emozionale: narrazioni virali, memi, teorie complottiste e versioni ipersemplificate della realtà che creano comunità forti ma, al contempo, epistemicamente fragili. Questo insieme di elementi si definisce come una sorta di terrapiattismo digitale, inteso come tendenza non relegata esclusivamente al mito della terra piatta, ma come processo di conversione della complessità in verità assolute e riduzioniste. La fiducia crescente nelle piattaforme e negli assistenti automatici depaupera l'esercizio critico, mentre l'estrema velocità della comunicazione favorisce schemi narrativi semplici e polarizzanti che causano la perdita della capacità di argomentazione, l'indebolimento del confronto e il contestuale aumento della vulnerabilità politica e sociale.

### **Sorveglianza, razionalizzazione e dominio**

I fenomeni sopracitati richiamano almeno tre tradizioni teoriche: la teoria critica della sorveglianza e del capitalismo comportamentale; la teoria della razionalizzazione e dell'alienazione e le analisi contemporanee su dominio algoritmico e «feudalesimo dei dati».

Shoshana Zuboff ha coniato il concetto di *surveillance capitalism* per descrivere la trasformazione dell'esperienza umana in materia prima, sfruttata per predire e manipolare comportamenti (Zuboff, 2019). Tale processo sposta potere decisionale dalle persone verso sistemi che raccolgono, profilano e monetizzano scelte, bisogni e aspettative, la conseguenza non è solo economica ma, come spiegato epistemica: quando il comportamento diventa predetto e indirizzato, l'incertezza su cui si fondano molte pratiche collettive si riduce artificialmente, minando l'autonomia deliberativa. (Zuboff, 2019). Weber ha descritto la modernità come un processo di razionalizzazione che costruisce una «gabbia d'acciaio» regolata da procedure, calcoli e burocrazia; oggi questa gabbia si attualizza attraverso architetture algoritmiche, non soltanto regole formali, ma codici che determinano visibilità, valore e

accesso alle informazioni (Weber, 1922). La conseguenza è duplice: a livello individuale si indebolisce la capacità di comprensione tecnico-critica; a livello sociale si consolidano gerarchie emergenti definite da chi controlla i metadati e l'economia dell'attenzione. Il concetto di *digital despotism* o dispotismo digitale attinge da questa linea critica e ne sviluppa una versione politica: la tecnologia non è neutra, ma può essere strumento di manipolazione ontologica e cognitiva quando incorpora pratiche che riducono la capacità degli individui di deliberare e auto-governarsi (Bhorat, 2023). La letteratura recente sottolinea che il pericolo non è solo la sorveglianza statale, ma forme ibride di dominio emergenti dall'intersezione tra aziende private, governance algoritmica e pratiche d'uso quotidiane. Infine, l'idea di *datafeudalism* o feudalesimo dei dati descrive una fase in cui piattaforme e ecosistemi digitali agiscono come signori feudali che detengono spazi informativi, instaurando rapporti di dipendenza e controllo nei confronti di utenti, istituzioni e mercati (Saura García, 2024). Tale modello spiega come la delega tecnologica si traduca nella cristallizzazione di rapporti di potere estremamente gerarchizzati, duraturi e non facilmente riformabili tramite la normazione tradizionale.

### **Definizione operativa di dispotismo digitale e terrapiattismo digitale**

Il *dispotismo digitale* rappresenta un processo, tecnico e sociale, che conduce alla delega massiva di funzioni cognitive, valutative e decisionali a sistemi tecnologici (algoritmi, piattaforme, assistenti), accompagnata dalla concentrazione di potere nelle mani di attori che, come spiegato, controllano dati, infrastrutture e regole di visibilità. Tale processo ha implicazioni ontologiche (chi possiede il «dato»), cognitive (chi decide cosa è rilevante) e politiche (chi definisce l'agenda) (Bhorat, 2023). Il *terrapiattismo digitale*, invece, si configura come un insieme di pratiche cognitive e discorsive che privilegiano spiegazioni semplici e definitive, spesso immuni alla verificabilità e alla falsificazione, che prosperano in ecosistemi informativi caratterizzati da una netta selezione algoritmica per engagement e da una forte polarizzazione emotiva. Si attua, insomma, una dinamica che non si limita alla «terra piatta» come dottrina, ma include ogni narrazione che converte ieraticamente la complessità in dogma semplificato (Landrum & Olshansky, 2019). Tali definizioni non si palesano come etichette moralistiche ma descrivono meccanismi profondamente interconnessi: la delega digitale alimenta la formazione di bolle epistemiche e di comunità che reagiscono alla complessità con miti e semplificazioni; le narrative semplici, a loro volta, facilitano, secondo un mero meccanicismo, l'adozione di tecnologie che promettono risposte rapide e rassicuranti.

### **Come la tecnologia favorisce delega e semplificazione**

Le piattaforme monetizzano l'attenzione e premiano contenuti che massimizzano il tempo di visualizzazione e la partecipazione emotiva. Tutto ciò crea incentivi sistematici a produrre contenuti divisivi ed estremamente semplificati, perché funzionali alla causa e, in questo modo, la moltiplicazione di segnali che detronizzano il ragionamento (like, share, view) converte il valore epistemico in valore di mercato. Gli algoritmi che filtrano e raccomandano informazioni sono spesso poco trasparenti: utenti e persino esperti di settore possono avere difficoltà a capire perché un contenuto è stato reso visibile: l'utente tende a fidarsi della piattaforma che «mostra» ciò che è rilevante, senza interrogarsi su criteri e metriche. La ricerca mostra che i sistemi di selezione possono

condurre i fruitori verso contenuti sempre più estremi o affini alla loro emotività, creando percorsi di radicalizzazione informativa. Gli assistenti digitali, i motori di ricerca e gli aggregatori offrono risposte rapide: l'utente medio, sotto pressione di tempo e sovraccarico informativo, tende a preferire risposte sintetiche e immediate. questo processo riduce l'uso di competenze critiche e la pratica argomentativa, favorendo la dipendenza da fonti immediate e dedicate piuttosto che dalla verifica (O'Connor, Weatherall, 2022).

### **Fenomenologia del terrapiattismo digitale**

Il fenomeno del terrapiattismo viene spesso derubricato come mera ignoranza, ma le ricerche sociologiche mostrano che la scelta di credere a narrazioni semplificate è legata a fattori psicologici, sociali e istituzionali: ricerca di senso, costruzione identitaria, sfiducia nelle istituzioni e bisogno di comunità (Arendt, 1961). in questo senso i movimenti terrapiattisti diventano esempi paradigmatici di come si costruisce fiducia sociale su basi non-epistemiche. La semplificazione risponde anche a bisogni emotivi: offrire spiegazioni nette riduce l'ansia dell'incertezza e rafforza legami di gruppo. Le piattaforme digitali forniscono gli strumenti per creare e mantenere comunità di significato alternative alle istituzioni tradizionali (chiesa, scuola, media), contribuendo così alla diffusione di visioni banalizzanti e impermeabili alla smentita. In questo senso la democrazia deliberativa richiede cittadini capaci di valutare informazioni, pesare argomenti e partecipare al dibattito pubblico, la delega digitale riduce queste capacità, spostando la responsabilità di filtrare e interpretare l'informazione a soggetti privati o a sistemi automatici.

La semplificazione, come detto, accentua la polarizzazione: storie nette favoriscono ampie identità contrapposte, rendendo più difficile il dialogo e aumentando la probabilità di conflitti simbolico-culturali. Quando il controllo dell'informazione è concentrato in poche piattaforme e la popolazione perde capacità critica, emergono spazi per forme morbide di autoritarismo: manipolazione dell'opinione pubblica, sospensione del dissenso e uso strategico di tecnologie per orientare comportamenti sociali, questa è la dimensione politica del dispotismo digitale. (Delfanti, Arvidsson, 2024).

### **Strategie di contrasto: educazione, progettazione e regolazione**

Contrastare dispotismo e terrapiattismo digitale richiede azioni su più livelli: educativo, progettuale e normativo, non basta l'insegnamento tecnico: è necessaria una alfabetizzazione che integri competenze critiche, etiche e affettive, cioè la capacità di riconoscere bias, capire incentivi delle piattaforme e gestire emotivamente la sovraesposizione informativa. Programmi scolastici e azioni di lifelong learning dovrebbero enfatizzare metodo scientifico, logica argomentativa e riconoscimento delle strategie persuasive. (Chourio-Acevedo et al. 2024). Le piattaforme possono essere ricalibrate per mitigare i meccanismi che favoriscono la semplificazione e la polarizzazione: maggiore trasparenza degli algoritmi, strumenti per ritardare o rendere meno automatico il flusso di raccomandazioni, ranking temporale alternativo, e opzioni che favoriscano la diversità informativa piuttosto che l'engagement puro. Interventi di «design for deliberation» possono riabilitare spazi pubblici digitali più trasparenti. Normative come il digital services act (dsa) nell'unione europea e altre iniziative di audit indipendente possono essere utili a riconoscere responsabilità delle piattaforme e a

garantire chiarezza sui meccanismi che strutturano l'informazione pubblica. Tuttavia, la regolazione deve essere accompagnata da strumenti di empowerment civico per evitare sostituzioni autoritarie della deliberazione popolare. (dibattiti su regolazione e antitrust). All'interno di questo iter analitico è importante non cadere in determinismi dogmatici: la tecnologia plasma ma non determina in modo univoco il comportamento sociale, molte realtà resistono alla semplificazione e alla delega: movimenti civici, giornalismo indipendente, comunità scientifiche palesano una robusta resilienza. Inoltre, molte piattaforme hanno introdotto correttivi e strumenti di moderazione, anche se spesso, purtroppo, insufficienti o contraddittori (Bhorat, 2023). Un altro limite consiste nella difficoltà di misurare la «regressione culturale» in termini empirici: occorrono indicatori efficaci per valutare la perdita di capacità argomentativa, la diminuzione della partecipazione e l'aumento della vulnerabilità che rende persuasive le narrative semplificate, in questo senso la ricerca dovrebbe sviluppare sistemi di misura interdisciplinari che integrino sinergicamente metodi qualitativi e quantitativi dedicati (O'Connor, Weatherall, 2022).

### **Conclusioni: tra pericolo e possibilità**

La riflessione sul dispotismo e sul terrapiattismo digitale mette in evidenza una dinamica complessa e, allo stesso tempo, preoccupante: la progressiva sostituzione della capacità critica individuale con forme di pensiero delegato e semplificato: in un ecosistema mediatico dominato da piattaforme centralizzate, il cittadino si trova sempre più spesso in una posizione di consumatore passivo di contenuti filtrati, per confermare pregiudizi e assecondare inclinazioni cognitive (Zuboff, 2019). Questa condizione, lungi dall'essere neutra, agisce come una vera e propria infrastruttura del potere informativo, dove la concentrazione delle leve comunicative nelle mani di pochi attori, i "feudatari" digitali, rende possibile un controllo sottile, continuo e spesso invisibile (Saura García, 2024). Il fenomeno del terrapiattismo digitale non rappresenta una curiosità marginale, ma è il sintomo di una regressione culturale che si manifesta nel rifiuto delle competenze esperte e nel privilegio fideistico accordato a narrazioni semplicistiche e complottiste. Il successo di queste narrazioni non risiede in un'ignoranza intrinseca e eziologica, ma deve essere ricondotto alla logica algoritmica che premia contenuti polarizzanti, emotivamente carichi e facili da assimilare (Delfanti, Ardvesson, 2024). In questo senso, il problema non è soltanto epistemologico ma strutturale: riguarda il modo in cui l'architettura tecnica dei media digitali interagisce con le predisposizioni cognitive e sociali degli utenti. Il rischio maggiore, dal punto di vista sociologico, non è solo quello di un'informazione distorta, ma di una cittadinanza ridotta a mere ritualità partecipative di facciata, dove la condivisione e il "mi piace" sostituiscono la deliberazione e il confronto pubblico (Villar-Onrubia et al., 2022). Il dispotismo digitale funziona perché si innesta su una cultura già abituata alla delega, in cui le scelte sono esternalizzate a meccanismi poco trasparenti: motori di ricerca, feed personalizzati, assistenti virtuali, che agiscono da mediatori totali della realtà percepita (Bhorat, 2023). La regressione culturale non è però un destino inevitabile. Ricerche recenti mostrano che interventi di alfabetizzazione digitale critica, soprattutto se introdotti nei contesti educativi fin dai primi livelli scolastici, possono aumentare la capacità degli individui di riconoscere manipolazioni, disinformazione e contenuti artificiali, come i deepfake (Geissler et al., 2025). Tuttavia, perché tali interventi siano efficaci, devono essere accompagnati da un rafforzamento delle competenze civiche e dalla promozione di una cultura della "lentezza informativa": la capacità di fermarsi, verificare, discutere e riformulare opinioni sulla

base di evidenze verificabili (Rowsell, 2025). In prospettiva, la sfida principale non è tanto quella di “combattere” i terrapiattisti digitali o di “demolire” le loro narrazioni, quanto di ricostruire spazi di senso condiviso e strumenti di comprensione complessa della realtà. Ciò implica una ridefinizione delle politiche educative, un ripensamento delle architetture tecnologiche e, soprattutto, un impegno collettivo a riappropriarsi della propria autonomia cognitiva, senza questa riappropriazione, il rischio è che la combinazione tra dispotismo digitale e terrapiattismo digitale diventi un dispositivo permanente di regressione culturale, capace di erodere le fondamenta stesse della cognizione antropica e della democrazia contemporanea.

## Bibliografia

- Arendt, H. (1961). *Between Past and Future*. New York: Penguin books.
- Bhorat, Z. (2023). Digital despotism and aristotle on the despotic master-slave relation. *philosophy & technology*, 36, articolo 77.
- Chourio-Acevedo, L., Köhler, J., Coscarelli, C., Gacitúa, D., Proaño-Ríos, V., González-Ibáñez, R. (2024). Information literacy development and assessment at school level: a systematic review of the literature. Arxiv preprint.
- Cinelli, M., et al. (2021). The echo chamber effect on social media. *Proceedings of the national academy of sciences*, 118(9).
- Daire, J. (2024). Challenging dominant perspectives of digital literacy in western higher education. Asclite publications.
- Delfanti, A., Arvidsson, A. (2024). *Introduzione ai media digitali*. Bologna: Il Mulino.
- Delfanti, A. (2019). machinic dispossession and augmented despotism: digital work in an amazon warehouse. *new media & society*, 23(1).
- Geissler, C., et al. (2025). Critical digital literacy in the age of ai. Cambridge: Cambridge University Press.
- Geissler, D., Robertson, C., Feuerriegel, S. (2025). Digital literacy interventions can boost humans in discerning deepfakes. Arxiv preprint.
- Ilomäki, L., et al. (2023). Critical digital literacies at school level: a systematic review. *Review of education*, Wiley.
- Kartika, R., Putri, A., Wita, G., Purnomo, S. (2025). 21st century sociology learning: integrating digital literacy, collaboration, and critical thinking. *Juwara: jurnal wawasan dan aksara*.
- O’ Connor C., Weatherall J. O., (2022). *The misinformation age*. New Haven: Yale university press.
- Lovink, G. (2019). *Sad by design: on platform nihilism*. London: Pluto press.
- Pariser, E. (2012). *The filter bubble: what the internet is hiding from you*. New York: Penguin press.
- Rowsell, J. (2025). *The comfort of screens: literacy in post-digital times*. Cambridge: Cambridge University Press.
- Saura García, C. (2024). The age of datafeudalism: from digital panopticon to synthetic democracy. *Philosophy & technology*, 37(3), articolo 98.
- Neubaum G., Chounta, I.-A., Gredel, E., Wiesche, D., (2025). A pandemic for the good of digital literacy? an empirical investigation of newly improved digital skills during covid-19 lockdowns. Arxiv preprint.

- Van Dijck, J., Poell, T., De Waal, M. (2018). *The platform society: public values in a connective world*. Oxford: Oxford University Press.
- Villar-Onrubia, D., Morini, L., Marín, V. I., Nascimbeni, F. (2022). Critical digital literacy as a key for (post)digital citizenship: an international review of teacher competence frameworks. *Journal of e-learning and knowledge society*.
- Zuboff, S. (2019). *The age of surveillance capitalism: the fight for a human future at the new frontier of power*. New York: PublicAffairs.



# AI e competenze umane: verso l'intelligenza "centaurica" – di Andrea Laudadio, Head of TIM Academy & Development

*L'avvento dell'intelligenza artificiale generativa provoca una metamorfosi delle competenze professionali. Nasce l'esigenza di nuove alfabetizzazioni digitali mentre le competenze tradizionali evolvono verso forme collaborative con sistemi AI avanzati*

## Abstract

L'Intelligenza Artificiale Generativa (GenAI) sta provocando una metamorfosi nel dominio delle competenze professionali e cognitive. Questo articolo analizza la bidimensionalità di tale trasformazione: da un lato, l'esigenza di sviluppare nuove competenze per interagire efficacemente con la GenAI; dall'altro, come la GenAI stessa riconfiguri le competenze esistenti. Attraverso una revisione della letteratura recente, vengono esplorate le dimensioni dell'AI literacy, del prompt engineering e dell'intelligenza "centaurica" - la sinergia tra intelligenza umana e artificiale. La rassegna si conclude delineando una visione evolutiva della relazione uomo-macchina e identificando direttive prioritarie per la ricerca futura.

## 1. Introduzione: Un'era di trasformazione delle competenze

L'avvento dell'Intelligenza Artificiale Generativa (GenAI), incarnata dai Large Language Model (LLM) come ChatGPT, sta innescando una trasformazione profonda in tutti i contesti lavorativi e sociali (D'Amora, 2024). Questa evoluzione tecnologica non rappresenta semplicemente l'aggiunta di un nuovo strumento all'arsenale professionale, ma una riconfigurazione radicale del concetto stesso di competenza.

La rapidità con cui questi sistemi si stanno diffondendo ha sollevato interrogativi fondamentali su quali siano le abilità necessarie per interagire efficacemente con essi e su come, simultaneamente, stiano riconfigurando competenze tradizionali considerate pilastri delle professioni contemporanee. Come evidenziato da Kasneci et al. (2023), stiamo assistendo a una vera metamorfosi nel dominio delle competenze umane, caratterizzata da una bidimensionalità: da un lato, l'esigenza di sviluppare nuove competenze per interagire criticamente con la GenAI; dall'altro, il modo in cui la GenAI stessa trasforma le competenze esistenti e può fungere da catalizzatore per lo sviluppo di nuove abilità.

Per comprendere la portata di questa trasformazione, basti considerare la previsione di Robin Li, co-fondatore e CEO di Baidu, secondo cui "tra dieci anni, metà dei lavori nel mondo coinvolgerà il prompt engineering, e coloro che non saranno in grado di scrivere prompt diventeranno obsoleti" (Korzyński et al., 2023, p. 26). Sebbene tale previsione possa apparire eccessiva, sottolinea l'importanza crescente delle competenze legate all'interazione con sistemi di AI.

Questo articolo si propone di esplorare la complessa relazione tra Intelligenza Artificiale Generativa e competenze umane, analizzando sia le "competenze per l'AI" sia l'"AI per le competenze". Attraverso una revisione della letteratura recente, verranno esaminate le dimensioni dell'AI literacy, l'emergere

del prompt engineering come competenza distintiva, e infine la riconfigurazione di competenze esistenti alla luce di questa nuova realtà tecno-sociale.

## 2. Competenze per l'AI: Verso una nuova alfabetizzazione

### 2.1 AI Literacy: Le fondamenta dell'interazione con l'Intelligenza Artificiale

L'interazione efficace con i sistemi di GenAI richiede un insieme di competenze specifiche che trascendono la tradizionale alfabetizzazione digitale. Federiakin et al. (2024) definiscono l'AI literacy come "la capacità di comprendere, valutare criticamente e utilizzare efficacemente i sistemi di AI, inclusa la consapevolezza delle loro capacità, limitazioni e implicazioni sociali" (p. 5). Questa definizione evidenzia la natura multidimensionale di questa nuova forma di alfabetizzazione, che integra aspetti tecnici, critici, pratici ed etici.

Le componenti dell'AI literacy possono essere articolate in quattro dimensioni fondamentali:

1. **Comprensione concettuale:** conoscenza dei principi basilari dell'AI, delle sue architetture e dei suoi limiti intrinseci
2. **Valutazione critica:** capacità di analizzare l'affidabilità degli output dell'AI, identificarne i bias e contestualizzarne le risposte
3. **Utilizzo efficace:** abilità di impiegare l'AI come strumento per risolvere problemi complessi e automatizzare processi
4. **Consapevolezza etica:** comprensione delle implicazioni sociali, legali ed etiche dell'uso dell'AI nei diversi contesti

Una definizione di AI Literacy la troviamo nell'AI Act (Art. 3): Le competenze, le conoscenze e la comprensione che consentono ai fornitori, ai deployer e alle persone interessate, tenendo conto dei loro rispettivi diritti e obblighi nel contesto del presente regolamento, di procedere a una diffusione informata dei sistemi di IA, nonché di acquisire consapevolezza in merito alle opportunità e ai rischi dell'IA e ai possibili danni che essa può causare (Gui e Galasso, 2025).

In questo quadro, il pensiero critico emerge come una competenza trasversale essenziale per navigare l'ecosistema dell'AI generativa. Broadfoot e Rockey (2025) introducono il concetto di "*connoisseurship*" come una nuova dimensione della valutazione educativa nell'era dell'AI, definendola come "la capacità di utilizzare efficacemente e valutare criticamente i contenuti generati dall'AI" (p. 282). Questa competenza implica non solo la capacità di interagire con l'AI, ma anche di discernere la qualità, l'accuratezza e l'utilità dei suoi output.

Holzmann et al. (2025) propongono un approccio interessante per lo sviluppo del pensiero critico basato proprio sulle limitazioni della GenAI: esaminando criticamente gli errori fattuali e le carenze nei contenuti generati dall'AI, gli individui possono sviluppare una comprensione più sofisticata dei processi di valutazione delle informazioni e verifica delle fonti. Questo approccio trasforma le imperfezioni dell'AI in opportunità pedagogiche per affinare il pensiero critico.

### 2.2 Prompt Engineering: La competenza emergente dell'era GenAI

Al centro delle nuove competenze richieste per interagire efficacemente con la GenAI si colloca il prompt engineering, definito come "il processo di progettare, elaborare e perfezionare input o domande contestualmente appropriate per suscitare tipi specifici di risposte o comportamenti da un modello linguistico di AI" (Bozkurt & Sharma, 2023) in modo coerente con le risposte desiderate (Gibreel & Arpacı, 2025).

Korzyński et al. (2023) concettualizzano il prompt engineering come una nuova competenza digitale, analizzandola attraverso il framework DigComp dell'Unione Europea. Gli autori sottolineano come questa competenza si sovrapponga a diverse aree del framework, descrivendola come "una pratica incentrata sul linguaggio umano in cui gli individui creano prompt utilizzando elementi di testo unici e separati, noti come token distinti" (p. 26).

Questa abilità emergente sta rapidamente acquisendo rilevanza nel panorama professionale contemporaneo, tanto da essere considerata un aspetto critico dell'alfabetizzazione AI (Walter, 2024) e riconosciuta come il "lavoro del futuro" numero uno a livello globale (Whiting, 2023). La sua importanza deriva dalla natura stessa dei modelli generativi, il cui output è fortemente condizionato dalla qualità e dalla struttura dell'input fornito.

Federiakin et al. (2024) approfondiscono ulteriormente la natura del prompt engineering, definendolo non come una semplice abilità tecnica, ma come "una complessa competenza cognitiva che combina elementi di pensiero critico, creatività e conoscenza del dominio" (p. 7). Questa caratterizzazione evidenzia la natura interdisciplinare del prompt engineering, che richiede sia competenze linguistiche sia una comprensione profonda del contesto applicativo.

La capacità di formulare input precisi, contestualizzati e iterativamente raffinati è fondamentale per guidare i modelli di GenAI verso gli output desiderati (White et al., 2023). Data la possibilità che la GenAI produca informazioni errate, incomplete o viziante da bias (Bender et al., 2021), è cruciale anche la capacità di analizzare criticamente i risultati, verificandone l'accuratezza, la coerenza e l'affidabilità.

Un contributo significativo alla misurazione delle competenze di prompting è stato fornito da Gibreel e Arpacı (2025), che hanno validato la PECS (Prompt Engineering Competence Scale). Questo strumento, composto da 9 item misurati attraverso una scala Likert a 5 punti, valuta diverse dimensioni delle competenze di prompting, inclusa la capacità di strutturare compiti complessi in prompt sequenziali e di valutare criticamente le risposte generate dall'AI. Gli indici psicometrici ottenuti confermano l'alta affidabilità e la robustezza dello strumento, offrendo una base metodologica per ulteriori ricerche in questo ambito.

Nel contesto educativo, Denny et al. (2024) propongono un approccio integrato per sviluppare simultaneamente competenze di comprensione del codice e di prompting, evidenziando come "con la diffusione dei modelli linguistici di grandi dimensioni (LLM), queste competenze fondamentali siano più importanti che mai, data la crescente necessità di comprendere e valutare il codice generato dai modelli" (p. 283). Gli autori sottolineano l'importanza di sviluppare competenze completamente nuove, come la capacità di formulare prompt chiari che possano guidare un LLM verso la generazione del codice desiderato.

### **3. L'AI per le competenze: Riconfigurazione delle abilità tradizionali**

---

### **3.1 Trasformazione delle competenze fondamentali**

L'avvento della GenAI sta provocando una profonda ridefinizione delle competenze umane fondamentali, a partire dalla scrittura. Questo processo, tradizionalmente considerato un'attività puramente umana, si sta evolvendo verso una collaborazione uomo-macchina, dove il focus si sposta dalla produzione di contenuto alla revisione, cura e direzione strategica.

Zamfirescu-Pereira et al. (2023) osservano come gli utenti stiano sviluppando nuove competenze di "meta-scrittura": la capacità di valutare criticamente, modificare e perfezionare il testo generato dall'AI, piuttosto che crearlo interamente da zero. Questo spostamento implica una riconfigurazione della competenza di scrittura, che diventa meno incentrata sulla produzione meccanica di testo e più orientata verso attività di supervisione e controllo qualitativo.

Parallelamente, assistiamo a una ridefinizione del concetto stesso di competenza linguistica o "proficiency". Se tradizionalmente questa era misurata in termini di padronanza grammaticale e lessicale, oggi si sta evolvendo verso la capacità di interagire efficacemente con sistemi AI attraverso il linguaggio naturale. Mishra et al. (2022) evidenziano come la formulazione di prompt efficaci richieda una sofisticata competenza linguistica, ma di natura diversa rispetto a quella tradizionale. Questo solleva interrogativi su come valutare la competenza linguistica in contesti professionali ed educativi dove l'AI è sempre più presente come co-autore o assistente.

Le abilità metacognitive – la capacità di riflettere sui propri processi di pensiero – subiscono anch'esse trasformazioni significative. Wu et al. (2022) descrivono come l'interazione con l'AI possa potenziare la metacognizione attraverso l'esternalizzazione del pensiero nei prompt, rendendo esplicativi processi cognitivi altrimenti impliciti. Tuttavia, esiste anche il rischio di una delega eccessiva dei processi metacognitivi all'AI, con conseguente potenziale atrofia di queste capacità fondamentali.

Emerge quindi la necessità di sviluppare una "meta-metacognizione": la consapevolezza di come l'AI influenza i propri processi cognitivi e la capacità di gestire strategicamente questa influenza. L'affidamento di compiti cognitivi alla GenAI può infatti indurre un "alleggerimento cognitivo" (cognitive offloading), un fenomeno già osservato con altre tecnologie (Sparrow et al., 2011; Carr, 2010) ma che la GenAI potrebbe amplificare. Se da un lato questo può liberare risorse cognitive per attività a più alto valore aggiunto, dall'altro necessita di un attento bilanciamento per evitare l'atrofia di abilità cognitive fondamentali.

### **3.2 Ridefinizione delle competenze professionali**

Le trasformazioni indotte dalla GenAI hanno profonde implicazioni per l'identità professionale in numerosi settori. Professioni tradizionalmente basate su competenze ora parzialmente automatizzabili – come la scrittura, la traduzione o l'analisi di dati – stanno vivendo una ridefinizione della propria essenza.

Jarrahi (2018) osserva l'emergere di nuove identità professionali ibride uomo-AI, dove il valore aggiunto umano si sposta verso competenze di supervisione, direzione creativa e valutazione critica (Gallotti, 2024).

Questo processo di adattamento non è privo di sfide psicologiche, richiedendo una rinegoziazione del proprio ruolo e valore in un ecosistema professionale in rapida evoluzione.

La riconfigurazione dei ruoli professionali si manifesta attraverso l'automazione di task routinari e l'aumento delle capacità umane in compiti complessi, un fenomeno analizzato da Autor (2015) e confermato da recenti studi sul futuro del lavoro (WEF, 2023). Ciò impone un reskilling e upskilling continuo orientato verso competenze complementari alla GenAI, come il pensiero critico, la creatività, l'empatia e la gestione della complessità.

Tummala (2025) propone una prospettiva interessante, ridefinendo il concetto di collaborazione e considerando l'AI come membro del gruppo piuttosto che come mero strumento. Secondo l'autore, "l'integrazione dell'AI come membro del team richiede nuove competenze collaborative, tra cui la capacità di delegare in modo appropriato, interpretare gli output dell'AI e mantenere la supervisione umana" (p. 12). Questa visione, sebbene possa apparire antropomorfizzante, evidenzia un importante cambio di paradigma nella concettualizzazione delle relazioni uomo-macchina in ambito professionale.

### **3.3 Impatto sui processi educativi e formativi**

Il campo dell'educazione e della formazione rappresenta un ambito privilegiato per osservare l'impatto della GenAI sulle competenze umane. Broadfoot e Rockey (2025) evidenziano come l'AI generativa stia trasformando le funzioni sociali della valutazione educativa, introducendo due nuove dimensioni oltre alle tradizionali "quattro C" (competenza, competizione, contenuto e controllo): "credibilità", riguardante l'autenticità della performance dello studente in un'era di output generati dall'AI, e "connoisseurship", relativa alla capacità di utilizzare efficacemente e valutare criticamente i contenuti generati dall'AI (p. 281).

McDonald et al. (2025) analizzano come le istituzioni di istruzione superiore stiano adattando le loro politiche e pratiche di valutazione in risposta all'AI generativa, rilevando una tendenza ad "abbracciare l'AI generativa fino al punto di approvare una revisione degli approcci pedagogici, senza considerare le implicazioni pedagogiche a lungo termine, l'investimento di tempo di istruttori e studenti, i cambiamenti nelle nozioni di proprietà, e con meno della metà che discute le implicazioni sulla privacy e l'etica dell'uso dell'AI generativa" (p. 2).

Tang et al. (2024) propongono un approccio dialogico per trasformare la valutazione educativa con l'AI generativa. Basandosi sul concetto di eteroglossia di Bakhtin, lo studio concettualizza l'AI generativa non come un fornitore definitivo di conoscenza, ma come un agente dialogico che facilita la collaborazione e la co-costruzione della conoscenza tra gli studenti. Questo approccio ridefinisce il concetto di competenza come capacità di partecipare attivamente alla co-costruzione della conoscenza attraverso il dialogo con l'AI, piuttosto che come mera acquisizione di informazioni.

Lu et al. (2024) esplorano gli effetti dell'AI generativa sullo sviluppo professionale degli insegnanti, evidenziando come l'AI stia cambiando il concetto di competenza docente. Nel loro studio, l'utilizzo della GenAI ha prodotto significativi miglioramenti nei livelli di competenza, suggerendo che l'AI generativa possa potenziare lo sviluppo di competenze di ordine superiore negli insegnanti, ridefinendo il concetto stesso di competenza docente.

Risultati simili sono stati osservati negli studenti. Chan et al. (2024) e Maphoto et al. (2024) hanno investigato l'impatto del feedback generato dalla GenAI sul miglioramento delle capacità di scrittura di saggi, dimostrando che gli studenti che hanno ricevuto feedback dall'AI hanno ottenuto miglioramenti significativi nella qualità dei loro elaborati. Inoltre, i risultati qualitativi hanno rivelato livelli più elevati di coinvolgimento, maggiore motivazione e risposte emotive miste al processo di feedback.

È importante, tuttavia, sottolineare che un uso passivo della GenAI può ostacolare lo sviluppo di competenze autentiche, favorendo la passività cognitiva (Adesope et al., 2017) e la perdita della "voce" personale (Montalvo, 2023). Questo evidenzia la necessità di un approccio equilibrato che integri l'AI nei processi educativi preservando al contempo lo sviluppo di competenze fondamentali e l'autonomia cognitiva degli studenti.

#### **4. Verso un'intelligenza "centaurica": La co-evoluzione uomo-AI**

La relazione tra Intelligenza Artificiale e competenze umane non dovrebbe essere concettualizzata in termini meramente oggettivi o in un rapporto di sostituzione, ma piuttosto come un'evoluzione simbiotica che porta all'emergere di nuove forme di intelligenza ibrida.

Un paradigma illuminante proviene dal mondo degli scacchi, dove Garry Kasparov, dopo la storica sconfitta contro DeepBlue, sviluppò il concetto di "centauri": esseri umani assistiti dai computer. Questa metafora, che deriva dalla creatura mitologica metà umana e metà equina, rappresenta efficacemente la fusione tra intelligenza umana e artificiale. La combinazione uomo-macchina, negli scacchi, si è dimostrata superiore sia ai soli esseri umani (anche di livello eccellente) sia ai soli computer, evidenziando il potenziale sinergico di questa integrazione.

Questa prospettiva suggerisce di concepire la GenAI non come uno strumento isolato o come un sostituto delle competenze umane, ma come un'estensione cognitiva, un "ponte" che amplifica le capacità umane e facilita l'interconnessione con il mondo circostante. Seguendo la metafora di Kasparov, la strada dell'"intelligenza centaurica" appare come la più promettente da percorrere, immaginando le competenze della GenAI come estensioni delle nostre capacità cognitive e riposizionando le competenze umane verso funzioni di supervisione, verifica, controllo, direzione, immaginazione e valorizzazione.

Questo paradigma centaurico implica una ridefinizione delle competenze umane non in opposizione all'AI, ma in complementarità con essa. La sfida diventa quindi quella di identificare e sviluppare quelle competenze distintamente umane che, integrate con le capacità computazionali e generative dell'AI, producono risultati superiori rispetto a entrambe le intelligenze operanti separatamente. Una integrazione centaurica aiuterebbe anche ad uscire dal problema, tautologico, dell'inseguimento continuo delle competenze d'uso e dell'effetto dell'uso sulle competenze.

In questa prospettiva, le competenze di prompt engineering rappresentano solo il primo livello di interazione, una sorta di interfaccia di base tra le due intelligenze. Il vero potenziale risiede nelle "competenze centauriche" emergenti: la capacità di orchestrare sinergie cognitive, di alternare fluidamente processi computazionali e intuitivi, di integrare la precisione algoritmica con l'immaginazione creativa.

## **5. Conclusioni e prospettive future**

L'avvento dell'Intelligenza Artificiale Generativa sta provocando una ridefinizione profonda del concetto di competenza, caratterizzata da una bidimensionalità: da un lato, la necessità di sviluppare nuove competenze per interagire efficacemente con questi sistemi; dall'altro, la trasformazione delle competenze esistenti sotto l'influenza pervasiva di questa tecnologia.

L'alfabetizzazione AI e il prompt engineering emergono come competenze fondamentali per navigare efficacemente l'ecosistema dell'AI generativa, richiedendo una combinazione di comprensione tecnica, pensiero critico, creatività e consapevolezza etica. Simultaneamente, competenze tradizionali come la scrittura, l'analisi linguistica e le abilità metacognitive stanno subendo una profonda trasformazione, evolvendo verso forme più integrate con i processi computazionali.

Il paradigma dell'"intelligenza centaurica" offre una prospettiva promettente per concettualizzare questa co-evoluzione, suggerendo che il futuro non sarà caratterizzato dalla sostituzione delle competenze umane, ma dalla loro integrazione sinergica con le capacità dell'AI. In questo scenario, il valore umano si riposiziona verso competenze di supervisione, direzione creativa e valutazione critica, mentre le funzioni computazionali e generative vengono delegate ai sistemi di AI.

La ricerca futura dovrebbe continuare a esplorare questa interdipendenza, sviluppando e validando strumenti di valutazione che catturino sia le "competenze per l'AI" sia l'impatto dell'"AI per le competenze". Studi longitudinali sulla co-evoluzione uomo-GenAI potrebbero fornire insights preziosi sugli effetti a lungo termine di questa simbiosi cognitiva, mentre approcci pedagogici innovativi dovrebbero essere sviluppati per bilanciare efficacemente l'insegnamento di "competenze per l'AI" con l'uso dell'"AI per lo sviluppo delle competenze" fondamentali.

Inoltre, è essenziale analizzare come questa relazione bidirezionale influenzi l'equità e le disuguaglianze, approfondendo le implicazioni etiche di questa interdipendenza, specialmente nella governance della collaborazione uomo-AI. Infine, esplorare ulteriormente modelli di "intelligenza centaurica", dove la sinergia tra competenze umane e capacità dell'AI produce risultati superiori, rappresenta una direzione di ricerca particolarmente promettente alla luce delle capacità crescenti della GenAI.

In conclusione, la sfida non è resistere all'avanzata dell'AI o utilizzarla passivamente, ma evolversi con essa, sviluppando quelle competenze distintamente umane che, in sinergia con le capacità dell'AI, ci permetteranno di navigare efficacemente la complessità crescente del mondo contemporaneo.

## **Bibliografia**

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, 87(3), 659-701.

Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, 29(3), 3-30.

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.

Bozkurt, A., & Sharma, R. C. (2023). Navigating the complexities of artificial intelligence in education. *Asian Journal of Distance Education*, 18(1), 1-5.

Broadfoot, P., & Rockey, J. (2025). Generative AI and the social functions of educational assessment. *Oxford Review of Education*, 51(2), 281-300.

Carr, N. (2010). *The shallows: What the Internet is doing to our brains*. W. W. Norton & Company.

Chan, S., Lo, N., & Wong, A. (2024). Generative AI and Essay Writing: Impacts of Automated Feedback on Revision Performance and Engagement. *rEFLections*, 31(3), 1249-1267.

D'Amora, F. (2024, 14 ottobre ). Adattarsi all'IA per non farsi sostituire: ecco le competenze che servono. Agenda Digitale. <https://www.agendadigitale.eu/cultura-digitale/competenze-digitali/adattarsi-allia-per-non-farsi-sostituire-ecco-le-competenze-che-servono/>

Denny, P., Smith IV, D. H., Fowler, M., Prather, J., Becker, B. A., & Leinonen, J. (2024). Explaining Code with a Purpose: An Integrated Approach for Developing Code Comprehension and Prompting Skills. *ITiCSE 2024*, 283-289.

Federiakin, D. A., Semenova, N. K., & Nikolaev, A. S. (2024). Prompt engineering as a new 21st century skill. *Journal of Educational Technology*, 15(2), 1-12.

Fuchs, A., Passarella, A., & Conti, M. (2024). Optimizing Delegation in Collaborative Human-AI Hybrid Teams. *ACM Transactions on Autonomous and Adaptive Systems*, 19(4), 1-33.

Gallotti, C. (2024, 26 giugno ). I professionisti dell'IA: chi sono e quali competenze devono avere. Agenda Digitale. <https://www.agendadigitale.eu/cultura-digitale/competenze-digitali/i-professionisti-della-chi-sono-e-quali-competenze-devono-avere/>

Gibreel, O., & Arpacı, I. (2025). Development and validation of the prompt engineering competence scale (PECS). *Computers in Human Behavior*, 146(8), 1-15.

Gnambs, T., Stein, J.-P., Appel, M., Grieser, F., & Zinn, S. (2025). An economical measure of attitudes towards artificial intelligence in work, healthcare, and education (ATTARI-WHE). *Computers in Human Behavior: Artificial Humans*, 3, 100106.

Gui, A., & Galasso, G. (2025, 18 marzo). AI Act, le competenze di "AI Literacy" richieste e come svilupparle. Agenda Digitale. <https://www.agendadigitale.eu/cultura-digitale/competenze-digitali/competenze-ia-requisiti-e-strategie-per-la-conformita-allai-act/>

Holzmann, V., Sarid, A., & Golan, O. (2025). The ChatGPT fact check: Exploiting the limitations of generative AI to develop evidence-based critical thinking. *Journal of Educational Technology*, 16(1), 1-18.

- Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, 61(4), 577-586.
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274.
- Korzynski, P., Mazurek, G., Krzypkowska, P., & Kurasinski, A. (2023). Artificial intelligence prompt engineering as a new digital competence: Analysis of generative AI technologies such as ChatGPT. *Entrepreneurial Business and Economics Review*, 11(3), 25-37.
- Lo, N. (2023). The Art and Science of Prompt Engineering: A New Literacy in the Information Age. *Journal of Digital Literacy*, 8(2), 45-62.
- Loock, R., & Holt, B. (2024). Augmented Linguistic Analysis Skills: Machine Translation and Generative AI as Pedagogical Aids for Analyzing Complex English Compounds. *Technology in Language Teaching & Learning*, 6(3), 1-27.
- Lu, J., Zheng, R., Gong, Z., & Xu, H. (2024). Supporting Teachers' Professional Development With Generative AI: The Effects on Higher Order Thinking and Self-Efficacy. *IEEE Transactions on Learning Technologies*, 17, 1267-1280.
- Maphoto, K. B., Sevnarayan, K., Mohale, N. E., Suliman, Z., Ntsopi, T. J., & Mokoena, D. (2024). Advancing Students' Academic Excellence in Distance Education: Exploring the Potential of Generative AI Integration to Improve Academic Writing Skills. *Open Praxis*, 16(2), 142-159.
- McDonald, N., Arden, S., Bonsignore, E., Clegg, T., Froehlich, J., Guha, M. L., Hurst, A., Kang, L., Pauw, D., & Yip, J. (2025). Generative artificial intelligence in higher education. *Computers in Human Behavior: Artificial Humans*, 3, 100121.
- Meskó, B. (2023). The imperative of artificial intelligence literacy in medical education. *JMIR Medical Education*, 9, e45317.
- Mishra, S., Sturm, B. L., & Dixon, S. (2022). Prompt tuning for generative multimodal pretrained models. *arXiv preprint arXiv:2208.08940*.
- Montalvo, N. (2023). The voice in the machine: Why we need to humanize AI in education. *Educational Technology Research and Development*, 71(1), 335-339.
- Pollino, M. A., Powell, E. A., & McCormick, M. L. (2025). Generative AI and the public-speaking course: A critical communication pedagogy approach. *Communication Teacher*, 39(1), 71-77.
- Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*, 333(6043), 776-778.

Tang, K.-S., Cooper, G., Rappa, N., Cooper, M., Sims, C., & Nonis, K. (2024). A dialogic approach to transform teaching, learning & assessment with generative AI in secondary education: a proof of concept. *Pedagogies: An International Journal*, 19(3), 493-503.

Tummala, R. (2025). AI as a Team Member: Redefining Collaboration. *Journal of Leadership Studies*, 12(1), 8-22.

Walter, S. L. (2024). The significance of prompt engineering in the age of generative AI: A review. *AI Communications*, 37(1), 41-57.

WEF (World Economic Forum). (2023). *Future of Jobs Report 2023*. World Economic Forum.

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). Prompt engineering techniques for text-to-image generative models. *Journal of Visual Communication and Image Representation*, 92, 103735.

Whiting, K. (2023). These are the top 10 jobs of the future, according to new research. *World Economic Forum Agenda*. <https://www.weforum.org/agenda/2023/05/jobs-skills-ai-future-of-work-wef-23/>

Wu, J., Ouyang, W., Wang, X., & Lian, Z. (2022). Beyond prompting: Making pre-trained language models better zero-shot learners by clustering responses. *arXiv preprint arXiv:2205.05836*.

Zamfirescu-Pereira, J. D., Valencia, R., Baur, D., Wong, R. Y., Haque, A., Muller, M., & Yang, Q. (2023). Why Johnny can't prompt: How non-AI experts try and fail to design LLM prompts. *CHI Conference on Human Factors in Computing Systems*, 1-17.

## Sezione DIRITTO DEL DIGITALE

### Direttori scientifici

**Carlo Colapietro**, Università degli Studi Roma Tre

**Antonio Iannuzzi**, Università degli Studi Roma Tre

### Comitato editoriale

**Giovanna Pistorio**, Università degli Studi Roma Tre

**Simone Barbareschi**, Università degli Studi Roma Tre

**Andrea Giubilei**, Università degli Studi Roma Tre

### Membri del Comitato scientifico

**Adriana Apostoli**, Università degli Studi di Brescia

**Antonio D'Aloia**, Università degli Studi di Parma

**Giovanni D'Alessandro**, Università degli Studi "Niccolò Cusano" Roma

**Massimo Durante**, Università degli Studi di Torino

**Federica Fabrizzi**, Università degli Studi di Roma La Sapienza

**Roberto Flor**, Università degli Studi di Trento

**Elisabetta Frontoni**, Università degli Studi Roma Tre

**Nicola Lupo**, Università LUISS

**Mario Midiri**, Università LUMSA

**Giovanna Pistorio**, Università degli Studi Roma Tre

**Fabrizio Politi**, Università degli Studi dell'Aquila

**Giovanni Maria Riccio**, Università degli Studi di Salerno

**Giorgio Resta**, Università degli Studi Roma Tre

**Massimo Rubechi**, Università degli Studi di Urbino

**Simone Scagliarini**, Università degli Studi di Modena e Reggio Emilia

**Andrea Simoncini**, Università degli Studi di Firenze

## Cybersicurezza e competenze digitali, pilastri dei diritti – di **Carlo Colapietro**, Professore ordinario di diritto costituzionale e pubblico presso il dipartimento di giurisprudenza dell'università degli studi Roma Tre

*La cybersicurezza diventa elemento cardine per tutelare i diritti fondamentali nell'ecosistema digitale. Senza competenze digitali adeguate, cittadini e istituzioni restano vulnerabili agli attacchi informatici che minacciano libertà e servizi essenziali*

Il cyberspazio è oggi diventato essenziale per l'esercizio di diritti fondamentali come il diritto all'immagine, la riservatezza, la libertà di manifestazione del pensiero e di informazione, nonché il diritto alla salute, il diritto di agire e resistere in giudizio e alcuni diritti connessi alla partecipazione democratica. Il cyberspazio, inteso come l'insieme delle infrastrutture digitali e delle reti di comunicazione che permettono l'interazione tra individui, imprese e Istituzioni, si configura oggi come una dimensione imprescindibile dell'esperienza umana. Non si tratta più di un ambito separato dalla realtà quotidiana, ma di uno spazio integrato nella vita sociale, politica, economica e culturale. In questo contesto, l'accesso al cyberspazio assume un valore strategico e costituisce una condizione necessaria per il pieno esercizio di una vasta gamma di diritti fondamentali, tra cui la libertà di espressione, il diritto all'informazione, la partecipazione democratica, il diritto all'istruzione e alla salute. Garantire un accesso equo e sicuro a tale ambiente significa, quindi, promuovere l'inclusione digitale e rafforzare la cittadinanza attiva.

Quanto più il digitale pervade l'esistenza quotidiana degli individui e delle amministrazioni, tanto più emerge il tema di come gli Stati debbano regolare il cyberspazio e garantire all'interno di esso i diritti degli individui. Infatti, non assicurare a questi ultimi il godimento dei diritti in Internet significa determinarne l'esclusione dal contesto sociale (digitale), nonché dal godimento dei diritti fondamentali e dall'esercizio dei doveri costituzionali. È stato, infatti, sottolineato come la cittadinanza oggi abbia acquisito una dimensione (anche) digitale: «l'innovazione tecnologica ha infatti aperto la strada a una nuova fase della vicenda di questa istituzione. Da *materiale*, essa si è fatta, in un certo senso *virtuale*»<sup>12</sup>.

Quindi, considerata la rilevanza del digitale nella vita dei singoli individui, garantire la cybersicurezza – per questa intendendosi «l'insieme delle attività necessarie per proteggere la

---

<sup>12</sup> F. AMORETTI, E. GARGIULO, *Dall'appartenenza materiale all'appartenenza virtuale. La cittadinanza elettronica fra processi di costituzionalizzazione della rete e dinamiche di esclusione*, in *Pol. dir.*, 3, 2010, 354, secondo cui «l'esercizio di molti diritti fondamentali è stato svincolato dal rapporto diretto tra il cittadino e gli apparati istituzionali, divenendo mediato da strumenti informatici che, per la loro stessa natura, si collocano al di fuori degli spazi fisici entro cui il rapporto tra il primo e i secondi ha comunemente luogo. La cittadinanza elettronica, in virtù della sua natura immateriale – in quanto tale adatta ad aggirare gli ostacoli di natura materiale che spesso si frappongono tra il cittadino e i suoi bisogni formalmente tutelati – sembrerebbe offrire inedite opportunità di esercitare diritti civili e politici fino a ora rimasti sulla carta e, nondimeno, di rivendicare nuovi tipi di diritti».

rete e i sistemi informativi, gli utenti di tali sistemi e altre persone interessate dalle minacce informatiche»<sup>13</sup> – significa garantire i diritti fondamentali che vengono esercitati nella “realtà online”<sup>14</sup>.

A sua volta, garantire la cybersicurezza richiede un’organizzazione istituzionale in cui l’Unione europea è chiamata a svolgere un ruolo da protagonista, che implica non solo una rivendicazione della *leadership* normativa, ma l’utilizzo di strumenti normativi funzionali alla difesa dei diritti fondamentali dell’individuo<sup>15</sup>. In questo senso, l’Unione europea è spesso accusata di regolare troppo, a differenza delle altre grandi superpotenze – come soprattutto gli USA – che invece lasciano più liberi i colossi del digitale. Tuttavia, contro le degenerazioni anti-libertarie che potrebbero emergere nel cyberspazio<sup>16</sup>, il rimedio non può essere costituito dall’assenza di regole o dal lasciare in mano a pochi soggetti privati<sup>17</sup> il dominio del cyberspazio e la sua regolazione. La risposta dell’Unione europea – e, a valle, degli Stati membri – è stata salda, negli ultimi anni, nel trasportare anche nello spazio cibernetico il pieno rispetto dei principi dello Stato di diritto. Ciò è necessario in quanto il «malfunzionamento deliberato di un sistema», dovuto ad attacchi informatici, «[impedisce] lo svolgimento di funzioni pubbliche o l’erogazione di servizi essenziali, cagionando importanti disservizi in grado addirittura di ledere diritti costituzionalmente riconosciuti ai cittadini (salute, istruzione, assistenza, lavoro, sicurezza ecc.)»<sup>18</sup>.

In questo senso, cybersicurezza e alfabetizzazione digitale diventano entrambi elementi cardine della tutela dei diritti fondamentali nell’ecosistema digitale, posto che il legame tra loro è strutturale. La cybersicurezza si basa sull’adozione di tecnologie, protocolli e pratiche volte a proteggere dati, reti e sistemi digitali da accessi non autorizzati, attacchi informatici e manipolazioni malevoli. Tuttavia, anche i sistemi più avanzati diventano vulnerabili se gli utenti non sono adeguatamente formati e consapevoli dei rischi connessi all’uso delle tecnologie digitali.

L’Italia registra un significativo ritardo rispetto alla media UE in materia di competenze digitali: nel 2023 solo il 46% degli adulti possedeva competenze digitali almeno di base. Il quadro è aggravato da divari generazionali e territoriali: tra i giovani (16-24 anni) la quota sale

---

<sup>13</sup> Art. 2 (“Definizioni”), punto 1), del Regolamento (UE) 2019/881 del Parlamento Europeo e del Consiglio del 17 aprile 2019 relativo all’ENISA, l’Agenzia dell’Unione europea per la cibersicurezza, e alla certificazione della cibersicurezza per le tecnologie dell’informazione e della comunicazione, e che abroga il regolamento (UE) n. 526/2013 («regolamento sulla cibersicurezza»).

<sup>14</sup> C. LOTTA, *Governance della Rete, accesso a Internet e cybersicurezza*, Editoriale Scientifica, Napoli, 2024, 181 ss.

<sup>15</sup> G. FINOCCHIARO, *Sovranità digitale*, in *Dir. Pubbl.*, 3, 2022, 827.

<sup>16</sup> Tra l’altro, anacronistiche, oggi, si rivelano le dichiarazioni di indipendenza del cyberspazio che, nel secolo scorso, propugnavano un modello di Internet senza la presenza degli Stati: si veda, in particolare, J.P. BARLOW, *Dichiarazione d’indipendenza del Cyberspazio*, in *Duke Law & Technology Review*, 5-7, 2019.

<sup>17</sup> ... comunque chiamati a rispettare il contenuto essenziale dei diritti fondamentali (C. CARUSO, *I custodi di silicio. Protezione della democrazia e libertà di espressione nell’era dei social network*, in AA.VV., *Liber Amicorum per Pasquale Costanzo*, Vol. I, in *Consulta Online*, 2020, 166).

<sup>18</sup> E. LONGO, *La disciplina della cybersicurezza nell’Unione europea e in Italia*, in F. PIZZETTI, S. CALZOLAIO, A. IANNUZZI, E. LONGO, M. OROFINO, *La regolazione europea della società digitale*, Giappichelli, Torino, 2024, 207.

al 59%, mentre tra gli over 65 crolla al 19%, con le regioni del Sud ampiamente sotto la media UE (34%).

L'alfabetizzazione digitale – intesa come la capacità non solo tecnica ma anche critica di interagire in modo consapevole con gli strumenti digitali – rappresenta quindi una componente essenziale della sicurezza informatica. In altre parole, la cybersicurezza è tanto una questione tecnologica quanto culturale, dato che senza cittadini digitalmente alfabetizzati, le strategie di sicurezza rischiano di essere inefficaci o incomplete.

Nel presente contributo, dopo aver esaminato – quantomeno per cenni – la normativa europea in materia di cybersicurezza (e aver richiamato alcune disposizioni significative in materia di alfabetizzazione digitale), si intende quindi esplorare il legame tra cybersicurezza e alfabetizzazione digitale, non solo con lo scopo di evidenziare la necessità di sviluppare competenze digitali solide e trasversali per la prevenzione e la gestione degli attacchi informatici, ma anche per dimostrare come la cybersecurity assurga oggi a vero e proprio “metainteresse” nella tutela dei diritti fondamentali.

### **Cenni sulla normativa europea e nazionale in materia di cybersicurezza (e di alfabetizzazione digitale)**

Il tema della cybersicurezza riguarda da vicino l'Italia, che risulta particolarmente colpita: a partire dal 2022, l'Italia è stata bersaglio di attacchi informatici in misura sempre crescente. In particolare, nel 2024 si è registrata una crescita degli incidenti informatici del 27% rispetto al 2023; inoltre, sempre nel 2024, gli incidenti “critici” o “gravi” hanno costituito l’80% del totale<sup>19</sup>. Nel 2024, l’ACN ha gestito 756 eventi cyber ai danni di istituzioni pubbliche nazionali, in sensibile aumento rispetto ai 383 del 2023: ascrivibile almeno in parte alle modifiche all’impianto normativo, oltre che alla più ampia capacità del CSIRT Italia di rilevare eventi, incidenti e criticità<sup>20</sup>. Si è, pertanto, dinanzi ad una “guerra cibernetica diffusa”.

Da qui la necessità di coordinamento tra Istituzioni europee, Istituzioni nazionali e Autorità di settore per una gestione sinergica della cybersicurezza, il cui prodotto è costituito dalla normativa europea volta a garantire la sicurezza delle reti, dei sistemi informativi e dei dati che circolano in Internet.

---

<sup>19</sup> ASSOCIAZIONE ITALIANA PER LA SICUREZZA INFORMATICA, *Rapporto Clusit sulla sicurezza ICT in Italia*, 2025, 7, reperibile al sito internet <https://clusit.it/rapporto-clusit/>.

<sup>20</sup> Agenzia per la cybersicurezza nazionale, *Relazione annuale al Parlamento 2024*, p. 46. Nel 2024 tutte le attività operative dell’ACN hanno subito un notevole incremento rispetto all’anno precedente, indice di un generale aumento della minaccia cyber: in particolare, 1.979 sono stati gli eventi cyber gestiti (165 al mese) e 573 incidenti con impatto confermato (48 al mese).

A tal proposito, occorre far preliminarmente riferimento alla Direttiva NIS 2<sup>21</sup> che, abrogando la Direttiva NIS 1<sup>22</sup>, stabilisce un livello comune elevato di cybersicurezza nell'Unione. Si consideri anzitutto che la fonte europea della direttiva richiede un approccio sinergico degli Stati europei, che devono tutti cooperare per raggiungere gli scopi individuati dalla direttiva stessa<sup>23</sup>, ovvero rafforzare la sicurezza cibernetica a livello europeo aumentando la sicurezza delle infrastrutture tecnologiche e combattendo in maniera efficace i rischi causati dal cybercrime. La Direttiva NIS 2, quindi, è stata varata al fine di rendere uniforme, nello spazio europeo, un elevato livello di cibersicurezza per gli operatori di rete “essenziali e importanti”<sup>24</sup>, a carico dei quali sono previsti specifici obblighi in materia di protezione della rete e dei sistemi informativi. Tale direttiva è stata recepita nel nostro Paese con il decreto legislativo n. 138 del 2024<sup>25</sup>.

Ancora a livello europeo, il Regolamento (UE) 2019/881 (*Cybersecurity Act*)<sup>26</sup>, “allo scopo di garantire il buon funzionamento del mercato interno perseguiendo nel contempo un elevato livello di cibersicurezza, ciberresilienza e fiducia nell’ambito dell’Unione”, definisce gli obiettivi, i compiti e gli aspetti organizzativi relativi all’ENISA («Agenzia dell’Unione europea per la cibersicurezza»)<sup>27</sup> e un quadro per l’introduzione di sistemi europei di certificazione della cibersicurezza.

---

<sup>21</sup> Direttiva (UE) 2022/2555 del Parlamento europeo e del Consiglio del 14 dicembre 2022, relativa a misure per un livello comune elevato di cibersicurezza nell’Unione, recante modifica del regolamento (UE) n. 910/2014 e della direttiva UE 2018/1972 e che abroga la direttiva (UE) 2016/1148 (direttiva NIS 1).

<sup>22</sup> Direttiva (UE) 2016/1148 del Parlamento Europeo e del Consiglio del 6 luglio 2016, recante misure per un livello comune elevato di sicurezza delle reti e dei sistemi informativi nell’Unione, recepita dal decreto legislativo 18 maggio 2018, n. 65.

<sup>23</sup> Per una critica all’utilizzo dello strumento della direttiva in materia di regolazione della cybersicurezza v. M. BUFFA, *La Direttiva NIS II Cybersecurity in Europa: tra innovazione, formazione e diritto vivente*, in *Democrazie e Diritti Sociali*, 1, 2023, 48, secondo cui «[...] l’obiettivo dell’uniformità delle legislazioni nazionali avrebbe certamente potuto essere perseguito con migliore e maggiore efficacia grazie all’adozione di un regolamento (direttamente applicabile senza la necessità di una normativa nazionale di recepimento). Ciò a partire dalla necessità di considerare le ricadute importanti della cybersecurity in materia di diritti fondamentali dei cittadini dell’UE, nonché in relazione alla evidente e sempre maggiore interconnessione tra le infrastrutture critiche dei diversi Stati membri».

<sup>24</sup> Direttiva (Ue) 2022/2555, art. 3 (“Soggetti essenziali e importanti”). Sul punto v. F. BAVETTA, *Direttiva NIS 2: verso un innalzamento dei livelli di cibersicurezza a livello europeo*, in *Media Laws*, 3, 2023, 408 e F. CASAROSA, *L’armonizzazione degli obblighi di notifica: il DDL Cybersecurity verso la NIS 2*, in *Rivista italiana di informatica e diritto*, 1, 2024, 13.

<sup>25</sup> Decreto legislativo 4 settembre 2024, n. 138, di recepimento della direttiva (UE) 2022/2555, relativa a misure per un livello comune elevato di cibersicurezza nell’Unione, recante modifica del regolamento (UE) n. 910/2014 e della direttiva (UE) 2018/1972 e che abroga la direttiva (UE) 2016/1148.

<sup>26</sup> Regolamento (UE) 2019/881 del Parlamento europeo e del Consiglio del 17 aprile 2019, relativo all’ENISA, l’Agenzia dell’Unione europea per la cibersicurezza, e alla certificazione della cibersicurezza per le tecnologie dell’informazione e della comunicazione, e che abroga il regolamento (UE) n. 526/2013 («regolamento sulla cibersicurezza»).

<sup>27</sup> Sul ruolo dell’ENISA v. L. PREVITI, *Pubblici poteri e cybersicurezza: il lungo cammino verso un approccio collaborativo alla gestione del rischio informatico*, in *Federalismi.it*, 25, 2022, 73 ss. e S. CALZOLAIO, *Autorità indipendenti e di governo nella società digitale*, in F. PIZZETTI, S. CALZOLAIO, A. IANNUZZI, E. LONGO, M. OROFINO (a cura di), *La regolazione europea della società digitale*, cit., 90 ss.

Più di recente, si pensi al *Cyber Resilience Act*<sup>28</sup>, che garantisce la sicurezza di prodotti e servizi, e al *Cyber Solidarity Act*<sup>29</sup>, che migliora la capacità dell'Unione di reagire agli attacchi informatici. Come è stato evidenziato, il diritto europeo della cybersicurezza a volte identifica la resilienza con la cybersicurezza; altre volte qualifica la prima come dimensione della seconda. In ogni caso, a prescindere dal *nomen*, il diritto dell'UE intende garantire sistemi che siano in grado di resistere agli attacchi informatici, di adattarsi nel caso in cui siano stati attaccati e di ripristinare le proprie capacità operative nel minor tempo possibile<sup>30</sup>.

Al livello nazionale si consideri il decreto-legge n. 105 del 2019<sup>31</sup>, che ha istituito il perimetro nazionale di sicurezza cibernetica (PNSC), “al fine di assicurare un livello elevato di sicurezza delle reti, dei sistemi informativi e dei servizi informatici delle amministrazioni pubbliche, degli enti e degli operatori pubblici e privati aventi una sede nel territorio nazionale, da cui dipende l'esercizio di una funzione essenziale dello Stato, ovvero la prestazione di un servizio essenziale per il mantenimento di attività civili, sociali o economiche fondamentali per gli interessi dello Stato e dal cui malfunzionamento, interruzione, anche parziale, ovvero utilizzo improprio, possa derivare un pregiudizio per la sicurezza nazionale” (art. 1)<sup>32</sup>.

Successivamente, è stato adottato dapprima il DPCM n. 131 del 2020<sup>33</sup>, con il quale sono stati individuati i soggetti pubblici e privati rientranti nel perimetro cibernetico nazionale e definiti i criteri per la predisposizione delle reti, dei sistemi informativi e dei sistemi informatici, e poi il DPCM n. 81 del 2021<sup>34</sup>, relativo, in particolare, alle notifiche degli incidenti<sup>35</sup>.

---

<sup>28</sup> Regolamento (UE) 2024/2487 del Parlamento europeo e del Consiglio del 23 ottobre 2024, relativo ai requisiti orizzontali di cibersicurezza per i prodotti con elementi digitali e che modifica i regolamenti (UE) n. 168/2013 e (UE) 2019/2020 e la direttiva (UE) 2020/1828 (regolamento sulla ciberresilienza).

<sup>29</sup> Regolamento (UE) 2025/38 del Parlamento europeo e del Consiglio del 19 dicembre 2024, che stabilisce misure intese a rafforzare la solidarietà e le capacità dell'Unione e di rilevamento delle minacce e degli incidenti informatici e di preparazione e risposta agli stessi, e che modifica il regolamento (UE) 2021/694 (regolamento sulla cibersolidarietà).

<sup>30</sup> Cfr. P.G. CHIARA, R. BRIGHI, *La dimensione della “resilienza” nel diritto Ue della cybersicurezza*, in *Ragion pratica*, 2, 2024, 422 s.

<sup>31</sup> Decreto-legge 21 settembre 2019, n. 105, recante “Disposizioni urgenti in materia di perimetro di sicurezza nazionale cibernetica”, convertito, con modificazioni, dalla legge 18 novembre 2019, n. 133.

<sup>32</sup> Sul tema v. S. POLETTI, *La sicurezza cibernetica nazionale ed europea, alla luce della creazione del perimetro di sicurezza nazionale cibernetica*, in *Media Laws*, 2, 2023, 404 ss.

<sup>33</sup> Decreto del Presidente del Consiglio dei Ministri 30 luglio 2020, n. 131, recante “Regolamento in materia di perimetro di sicurezza nazionale cibernetica, ai sensi dell'art. 1, comma 2, del decreto-legge 21 settembre 2019, n. 105, convertito, con modificazioni, dalla legge 18 novembre 2019, n. 133”.

<sup>34</sup> Decreto del Presidente del Consiglio dei ministri 14 aprile 2021, n. 81, recante “Regolamento in materia di notifiche degli incidenti aventi impatto su reti, sistemi informativi e servizi informatici di cui all'art. 1, comma 2, lettera b), del decreto-legge 21 settembre 2019, n. 105, convertito, con modificazioni, dalla legge 18 novembre 2019, n. 133, e di misure volte a garantire elevati livelli di sicurezza”.

<sup>35</sup> Art. 1, comma 1, lett. h).

Quindi, con il decreto-legge n. 82 del 2021<sup>36</sup>, si è “ritenuto [...] di dover intervenire con urgenza al fine di ridefinire l’architettura italiana di cybersicurezza, prevedendo anche l’istituzione di un’apposita Agenzia per la cybersicurezza nazionale, per adeguarla all’evoluzione tecnologica, al contesto di minaccia proveniente dallo spazio cibernetico, nonché al quadro normativo europeo, e di dover raccordare, altresì, pure a tutela dell’unità giuridica dell’ordinamento, le disposizioni in materia di sicurezza delle reti, dei sistemi informativi, dei servizi informatici e delle comunicazioni elettroniche”<sup>37</sup>.

Con il decreto-legge in esame è stata attribuita in via esclusiva al Presidente del Consiglio dei Ministri un cospicuo numero di funzioni nell’ambito della cybersicurezza<sup>38</sup>. In particolare, egli ha la competenza dell’alta direzione e della responsabilità generale delle politiche di cybersicurezza; dell’adozione della strategia nazionale di cybersicurezza, sentito il Comitato interministeriale per la cybersicurezza (CIC); della nomina e della revoca del direttore generale e del vicedirettore generale dell’Agenzia per la cybersicurezza nazionale, previa deliberazione del Consiglio dei Ministri”<sup>39</sup>.

Nell’ambito del decreto-legge n. 82 del 2021 assume primaria importanza l’art. 5 con il quale viene istituita – colmando «una grave lacuna nell’architettura istituzionale nazionale in materia di cybersicurezza»<sup>40</sup> - l’Agenzia per la cybersicurezza nazionale (ACN) “a tutela degli interessi nazionali nel campo nella cybersicurezza”. Le funzioni dell’ACN includono la tutela degli interessi nazionali in ambito di cybersicurezza, la predisposizione della strategia nazionale e la qualificazione dei servizi *cloud* per la Pubblica amministrazione. L’Agenzia rappresenta, inoltre, il punto di contatto unico per le finalità del decreto NIS, potendo irrogare sanzioni agli operatori di servizi essenziali o ai fornitori di servizi digitali; assume le funzioni in materia di cybersicurezza del Ministero dello Sviluppo Economico, del Dipartimento delle Informazioni per la Sicurezza e della stessa AgId. L’istituzione dell’ACN è strumentale all’esercizio delle competenze che il d.l. n. 82/2021 assegna al Presidente del Consiglio dei Ministri – l’autorità al vertice dell’architettura della sicurezza cibernetica – al quale è attribuita l’alta direzione e la responsabilità generale delle “politiche di cybersicurezza”, ed a cui spetta

---

<sup>36</sup> Decreto-legge 14 giugno 2021, n. 82, recante “*Disposizioni urgenti in materia di cybersicurezza, definizione dell’architettura nazionale di cybersicurezza e istituzione dell’Agenzia per la cybersicurezza nazionale*”, convertito con modificazioni dalla legge 4 agosto 2021, n. 109. Sul decreto-legge in esame si veda F. SERINI, *La nuova architettura di cybersicurezza nazionale: note a prima lettura del decreto-legge n. 82 del 2021*, in *Federalismi.it*, 12, 2022, 241 ss.

<sup>37</sup> Così il Preambolo del decreto-legge 14 giugno 2021, n. 82. Sulle diverse connotazioni della minaccia cibernetica, v. M. MENSI, *La sicurezza cibernetica*, in M. MENSI, P. FALLETTO (a cura di), *Il diritto nel web*, Cedam, 2018, 282 ss., che distingue le minacce cibernetiche nelle seguenti tipologie: *cyber-crime*, *cyber-espionage*, *cyber-terrorism*, *cyber-warfare*.

<sup>38</sup> Sul tema v. L. MORONI, *La governance della cybersicurezza a livello interno ed europeo: un quadro intricato*, in *Federalismi.it*, 14, 2024, 192 ss. ed O. CARAMASCHI, *La cybersicurezza nazionale ai tempi della guerra (cibernetica): il ruolo degli organi parlamentari*, in *Osservatorio AIC*, 4, 2022, 76 ss.

<sup>39</sup> Art. 2 (“Competenze del Presidente del Consiglio dei Ministri”). Sul ruolo della Presidenza del Consiglio dei Ministri nell’ambito della cybersecurity v. T. F. GIUPPONI, *Il governo nazionale della cybersicurezza*, in *Quad. Cost.*, 2, 2024, 295 ss.

<sup>40</sup> L. PARONA, *L’istituzione dell’Agenzia per la cybersicurezza nazionale*, in *Giornale di diritto amministrativo*, 6, 2021, 711.

l'adozione della relativa strategia nazionale, nonché la nomina (e la revoca) dei vertici dell'Agenzia (Direttore e Vicedirettore)<sup>41</sup>.

Successivamente è stata approvata, più di recente, la legge n. 90 del 2024<sup>42</sup>, che non costituisce diretta attuazione della Direttiva NIS 2, piuttosto «ne anticipa alcune misure, senza tuttavia prevedere forme di raccordo con essa. L'intervento proposto risulta quindi asincrono, e rischia di produrre ulteriore incertezza»<sup>43</sup>.

La legge n. 90 del 2024 si presenta eterogenea, quanto ai contenuti, atteso che al Capo I prevede “*Disposizioni in materia di rafforzamento della cybersicurezza nazionale, di resilienza delle pubbliche amministrazioni e del settore finanziario, di personale e funzionamento dell'Agenzia per la cybersicurezza nazionale e degli organismi di informazione per la sicurezza nonché di contratti pubblici di beni e servizi informatici impiegati in un contesto connesso alla tutela degli interessi nazionali strategici*”. Il Capo II, invece, raccoglie le “*Disposizioni per la prevenzione e il contrasto dei reati informatici nonché in materia di coordinamenti degli interventi in caso di attacchi a sistemi informatici o telematici e di sicurezza delle banche dati in uso presso gli uffici giudiziari*”.

È stato rilevato che all'eterogeneità delle disposizioni si affianca un'altra criticità, dovuta alla presenza della consueta clausola di invarianza finanziaria, per cui «l'intero progetto non è credibile se non si individuano risorse adeguate»<sup>44</sup>. Ciò contrasta con la fondamentale esigenza di formazione sulla cibersicurezza, che, «parte di un deficit di conoscenze digitali dei cittadini italiani, deve essere assunta come un'emergenza democratica ed una questione costituzionalmente rilevante»<sup>45</sup>. Non solo: la legge n. 90 del 2024 pone l'accento sugli aspetti sanzionatori, ma non si occupa del profilo inerente all'aggiornamento delle competenze del personale della Pubblica amministrazione che quegli attacchi dovrebbe essere in grado di riconoscere e reprimere<sup>46</sup>.

Pur non essendo possibile, in questa sede, richiamare tutte le disposizioni e le iniziative dedicate allo sviluppo delle competenze digitali, appare comunque utile menzionare le principali, per testimoniare lo sforzo delle Istituzioni europee e nazionali in materia di alfabetizzazione digitale. È opportuno partire dall'art. 8 del Codice dell'amministrazione

---

<sup>41</sup> Sul punto v. A. VENANZONI, *L'ordine costituzionale della cybersecurity*, in *Forum Quad. Cost.*, 4, 2024, 62.

<sup>42</sup> Legge 28 giugno 2024, n. 90, recante “*Disposizioni in materia di rafforzamento della cybersicurezza nazionale e di reati informatici*”.

<sup>43</sup> Così M. PIETRANGELO, *Per un modello nazionale di cybersicurezza cooperativa e resilienza collaborativa*, in *Rivista italiana di informatica e diritto*, 1, 2024, 25. V. anche E. LONGO, *Audizione informale per il disegno di legge in materia di «Disposizioni in materia di rafforzamento della cybersicurezza nazionale e di reati informatici»* (AC 1717), in *Rivista italiana di informatica e diritto*, 1, 2024, 68.

<sup>44</sup> Cfr. ancora M. PIETRANGELO, *Per un modello nazionale di cybersicurezza cooperativa e resilienza collaborativa*, cit., 26 s., che quindi, propone di ancorare le misure di cui alla legge n. 90 del 2024 ai fondi destinati alla cybersicurezza istituiti con la legge n. 197 del 2022 (il “*Fondo per l'attuazione della Strategia nazionale di cybersicurezza*” e il “*Fondo per la gestione della cybersicurezza*”).

<sup>45</sup> A. IANNUZZI, *Considerazioni sul disegno di legge «Disposizioni in materia di rafforzamento della cybersicurezza nazionale e di reati»* (AC 1717), in *Rivista italiana di informatica e diritto*, 1, 2024, 60.

<sup>46</sup> E. LONGO, *Audizione informale*, cit., 68.

digitale, di cui al decreto legislativo 7 marzo 2005, n. 82. La disposizione (rubricata proprio “Alfabetizzazione informatica dei cittadini”) è stata peraltro modificata e ampliata con il decreto legislativo 26 agosto 2016, n. 179<sup>47</sup> e ora stabilisce che “lo Stato e i soggetti di cui all’articolo 2, comma 2<sup>48</sup>, promuovono iniziative volte a favorire la diffusione della cultura digitale tra i cittadini con particolare riguardo ai minori e alle categorie a rischio di esclusione, anche al fine di favorire lo sviluppo di competenze di informatica giuridica e l’utilizzo dei servizi digitali delle pubbliche amministrazioni con azioni specifiche e concrete, avvalendosi di un insieme di mezzi diversi fra i quali il servizio radiotelevisivo”<sup>49</sup>.

Anche a livello di Unione europea è stata riservata una certa attenzione al tema, a partire almeno dall’art. 33-bis, paragrafo 3, della direttiva (UE) 2018/1808, in base al quale “gli Stati membri promuovono lo sviluppo dell’alfabetizzazione mediatica e adottano misure a tal fine”. L’impegno è proseguito, a livello europeo, con il piano d’azione per l’istruzione digitale (2021-2027), un’iniziativa politica rinnovata dell’Unione europea che definisce una visione comune di un’istruzione digitale di alta qualità, inclusiva e accessibile in Europa e che punta a sostenere l’adeguamento dei sistemi di istruzione e formazione degli Stati membri all’era digitale. Lo sviluppo delle competenze digitali è uno degli obiettivi strategici dell’Unione, che mira a dotare almeno l’80% dei cittadini tra i 16 e i 74 anni di competenze digitali di base entro il 2030. A guidare questo processo è il quadro europeo DigComp 2.2, che individua cinque aree chiave – dalla sicurezza informatica, alla creazione di contenuti – ritenute essenziali per una piena partecipazione alla vita sociale e professionale.

Tornando al versante nazionale, nel 2020 si è adottata la Strategia Nazionale per le Competenze Digitali con l’obiettivo di eliminare il *gap* con gli altri Paesi europei, in termini generali di digitalizzazione e rispetto ai singoli assi di intervento, e di abbattere il *digital divide* tra varie aree del territorio nazionale. Il Piano nazionale di ripresa e resilienza ha poi, da ultimo, dedicato particolare attenzione al tema, attivando anche la misura “Rete dei servizi di facilitazione digitale”, con l’obiettivo di formare 2 milioni di cittadini entro il 2026 attraverso 3000 Punti Digitale Facile.

La rapida panoramica sulle fonti che disciplinano la cybersicurezza mostra come le normative hanno posto un forte accento sulla protezione delle infrastrutture critiche, sulla

---

<sup>47</sup> Recante “Modifiche ed integrazioni al Codice dell’amministrazione digitale, di cui al decreto legislativo 7 marzo 2005, n. 82, ai sensi dell’articolo 1 della legge 7 agosto 2015, n. 124, in materia di riorganizzazione delle amministrazioni pubbliche”.

<sup>48</sup> L’art. 2, comma 2 del CAD recita che: Le disposizioni del presente Codice si applicano:

a) alle pubbliche amministrazioni di cui all’[articolo 1, comma 2, del decreto legislativo 30 marzo 2001, n. 165](#), nel rispetto del riparto di competenza di cui all’[articolo 117 della Costituzione](#), ivi comprese le autorità di sistema portuale, nonché alle autorità amministrative indipendenti di garanzia, vigilanza e regolazione;

b) ai gestori di servizi pubblici, ivi comprese le società quotate, in relazione ai servizi di pubblico interesse;

c) alle società a controllo pubblico, come definite nel [decreto legislativo 19 agosto 2016, n. 175](#), escluse le società quotate di cui all’articolo 2, comma 1, lettera p), del medesimo decreto che non rientrino nella categoria di cui alla lettera b).

<sup>49</sup> La disposizione, prima della modifica del 2016, si limitava invece a stabilire che “Lo Stato promuove iniziative volte a favorire l’alfabetizzazione informatica dei cittadini con particolare riguardo alle categorie a rischio di esclusione, anche al fine di favorire l’utilizzo dei servizi telematici delle pubbliche amministrazioni”.

tutela dei dati personali e sulla repressione degli attacchi informatici. Come si è visto, anche l’alfabetizzazione digitale ha ricevuto una certa attenzione dal legislatore europeo e da quello nazionale. Eppure, l’alfabetizzazione digitale rappresenta un elemento chiave per la costruzione di una società resiliente e consapevole, in grado di riconoscere, prevenire e reagire efficacemente alle minacce informatiche. Queste lacune rischiano di indebolire la capacità del Paese di affrontare le sfide poste dalla digitalizzazione crescente, lasciando ampi spazi di vulnerabilità non tanto nelle tecnologie, ma nelle competenze degli utenti stessi.

Peraltro, sapersi muovere nel cyberspazio in sicurezza e in modo consapevole è ormai una necessità imprescindibile. È fondamentale, infatti, che ogni cittadino sia in grado di padroneggiare le nuove tecnologie per muoversi con competenza e tutela nel mondo digitale. In questo contesto, la scuola può e deve assumere un ruolo centrale, rappresentando il principale veicolo per diffondere conoscenze e competenze digitali fin dalla giovane età.

È incoraggiante, a tal proposito, il recente Protocollo d’intesa stipulato tra il Dipartimento per la trasformazione digitale (DTD) e l’Agenzia per la cybersicurezza nazionale (ACN), col quale si è dato vita a una nuova sinergia volta a potenziare le attività di sensibilizzazione e accrescimento delle competenze digitali sulla cybersicurezza e rafforzare la consapevolezza dei cittadini sulla materia<sup>50</sup>. L’accordo testimonia quantomeno della raggiunta consapevolezza, da parte delle Istituzioni italiane, sull’importanza del legame sempre più stretto che occorre instaurare tra cybersicurezza e competenze digitali.

### **Necessità di sviluppare competenze digitali per la prevenzione e la gestione degli attacchi informatici**

Dall’analisi normativa appare evidente come l’Unione europea e il legislatore statale abbiano inteso dar vita ad un vasto *corpus* normativo volto alla tutela delle infrastrutture di rete e dei dati che circolano in Internet, considerato che la cybersicurezza è funzionale all’esercizio dei diritti fondamentali: in assenza di dispositivi tecnologicamente avanzati atti a tutelare le reti e senza un idoneo strumentario normativo non può esistere la cybersicurezza. Tuttavia, si deve subito aggiungere che la creazione di un apparato – anzitutto normativo – volto a disciplinare la cybersicurezza è condizione necessaria, ma non sufficiente per assicurare la tutela dei diritti fondamentali dell’individuo nell’ecosistema digitale. Infatti, sia le direttive ed i regolamenti europei che le leggi necessitano di soggetti che siano in grado di

---

<sup>50</sup> L’iniziativa, che si inserisce nel contesto del PNRR, coinvolge gli oltre 3.300 “Punti Digitale Facile”, attivati nell’ambito della misura 1.7.2 del PNRR e gli Operatori volontari del Servizio Civile Digitale (misura 1.7.1), insieme alle oltre 280 organizzazioni della Coalizione Nazionale per le competenze digitali, parte integrante del programma strategico “Repubblica Digitale”. Le attività di formazione prevedono anche materiali formativi di supporto ai Facilitatori, Operatori Volontari e organizzazioni della Coalizione nell’erogazione di corsi di formazione, workshop, ed eventi di sensibilizzazione sull’uso sicuro e consapevole degli strumenti digitali. La campagna di sensibilizzazione coinvolgerà anche i canali social del DTD e di ACN, con l’obiettivo di promuovere le attività svolte sul territorio e di diffondere strumenti e buone pratiche.

avvalersi delle tecnologie sulla cybersicurezza. In altre parole, è necessario che pubbliche istituzioni e soggetti privati siano supportati nell'acquisizione (o nel miglioramento) delle competenze digitali.

Pertanto, la cybersicurezza richiede non solo adeguamenti normativi, ma anche un profondo cambio di impostazione culturale. Così come, nel corso degli ultimi dieci anni, si è assistito all'affermarsi di una cultura della protezione dei dati personali, incentivata anche da strumenti normativi come il GDPR, oggi è necessario che vi sia uno slancio per favorire la cultura della cybersicurezza.

Il tema in esame, invero, ha molto a che fare con l'alfabetizzazione digitale volta a ridurre il c.d. divario digitale (*digital divide*), inteso quale discriminazione tra coloro che sono in grado di utilizzare gli strumenti informatici e coloro che, per ragioni economiche, culturali, generazionali, non sono in grado di avvalersene<sup>51</sup>. Il tema della alfabetizzazione digitale – preso in considerazione nel PNRR<sup>52</sup> – ha infatti a che vedere con il concetto di “cittadinanza digitale” perché, come visto, essa oggi si sviluppa anche (forse, soprattutto) in una dimensione digitale. Non essere in grado di utilizzare gli strumenti informatici può comportare – e di fatto comporta – il rischio di una emarginazione dalla società e nell'esclusione dall'esercizio di molti diritti fondamentali.

Ebbene, è necessario operare una differenziazione tra alfabetizzazione digitale in senso stretto (rientrante, cioè, sotto la tutela dell'art. 34 della Costituzione e rivolta a coloro che sono in età scolastica<sup>53</sup>) e diritto all’“inclusione sociale digitale”. Quest’ultimo trova, infatti, il suo fondamento costituzionale nell'art. 3, comma 2, Cost. ed è stato configurato dal legislatore come un diritto sociale per il cui inveramento sono state stanziate risorse economiche volte a garantirne l’attuazione<sup>54</sup>.

Quindi, sia l'alfabetizzazione digitale che il diritto all'inclusione sociale digitale devono essere orientati alla implementazione di conoscenze digitali in materia di cybersicurezza, soprattutto se si considera che la normativa in materia di cybersecurity riguarda quasi esclusivamente le Pubbliche Amministrazioni e le imprese (soprattutto le piccole e medie imprese). Della cybersecurity, tuttavia, deve aver cognizione anche il singolo cittadino, perché possa tutelarsi dai rischi che quotidianamente si corrono nella navigazione in Internet. Solo in

---

<sup>51</sup> L. NANNIPIERI, *Costituzione e nuove tecnologie: profili costituzionali dell'accesso a Internet*, in *Rivista dell'Associazione "Gruppo di Pisa"*, Secondo seminario annuale del “Gruppo di Pisa” con i dottorandi delle discipline giuridiche su “Lo studio delle fonti del diritto e dei diritti fondamentali in alcune ricerche dottorali”, Università degli Studi Roma Tre, 20 settembre 2013, 3.

<sup>52</sup> Il Piano Nazionale di Ripresa e Resilienza (PNRR) alla Missione 1 (“Digitalizzazione, innovazione, competitività, cultura e turismo”), componente 1 (“Digitalizzazione, innovazione e sicurezza nella PA”), investimento 1.7 (“Competenze digitali di base”) prevede interventi miranti a supportare le fasce della popolazione che, più delle altre, potrebbero subire le conseguenze negative del *digital divide*. In particolare, la missione in esame mira a rafforzare il «network territoriale di supporto digitale» e il «Servizio Civile Digitale, attraverso il reclutamento di diverse migliaia di giovani che aiutino circa un milione di utenti ad acquisire competenze digitali di base». Il programma “Servizio Civile Digitale”, sub investimento 1.7.1., per il quale sono stanziati 60 milioni di euro, si inserisce nel quadro della “Strategia nazionale per le competenze digitali” e del connesso Piano Operativo, entrambi sviluppati nell’ambito del progetto “Repubblica digitale”.

<sup>53</sup> Sul punto si consideri il Piano nazionale per la scuola digitale, istituito dall'art. 1, comma 56, della legge 13 luglio 2015, n. 107.

<sup>54</sup> Come rileva C. LOTTA, *Governance della Rete*, cit., 157 ss.

tal modo si formano individui in grado di riconoscere e gestire minacce informatiche che possono minare i loro diritti e costituire un impedimento all'adempimento dei propri doveri.

In un contesto digitale sempre più interconnesso e vulnerabile, la diffusione di competenze digitali si impone, quindi, come una priorità strategica per garantire la sicurezza dei sistemi informativi e la protezione dei dati personali e istituzionali. Gli attacchi informatici, per loro natura mutevoli e sofisticati, richiedono risposte che vadano oltre le sole misure tecnologiche: è necessario che i cittadini siano in grado di comprendere, anticipare e gestire le minacce informatiche in modo consapevole e proattivo.

Lo sviluppo di tali competenze non deve essere limitato agli specialisti, ma deve riguardare l'intera popolazione, attraverso percorsi di alfabetizzazione digitale diffusa e continua, integrati nei sistemi educativi, nella formazione professionale e nelle politiche pubbliche. In questo quadro, le competenze digitali rappresentano non solo una leva di innovazione, ma anche un presidio essenziale di sicurezza e di esercizio dei diritti nella sfera digitale.

Infine, sul versante dell'alfabetizzazione digitale in senso stretto – rientrante, come si è detto, nel diritto all'istruzione di cui all'art. 34 Cost.<sup>55</sup> – è auspicabile che il legislatore implementi un percorso strutturato di educazione digitale che comprenda in modo esplicito e approfondito anche la formazione alla cybersicurezza. Incorporare, infatti, l'educazione alla cybersicurezza nel percorso scolastico significa fornire ai giovani non solo gli strumenti tecnici di base, ma anche farli giungere a comportamenti responsabili, fondamentali per prevenire i rischi più diffusi.

La percentuale di persone prive di competenze digitali adeguate resta ancora troppo elevata, rappresentando un ostacolo significativo all'inclusione piena nella società digitale<sup>56</sup>.

Per colmare queste carenze, un primo significativo passo è rappresentato dalla legge 20 agosto 2019, n. 92 “*Introduzione dell'insegnamento scolastico dell'educazione civica*”, che all'art. 5 ha appunto introdotto l'educazione digitale nell'ambito dell'insegnamento trasversale dell'educazione civica, le cui modalità di attuazione sono state definite con le Linee guida ministeriali emanate con il recente D.M. 7 settembre 2024, n. 183.

Ma non basta; parallelamente si tratta anche di garantire agli istituti scolastici dotazioni tecnologiche moderne e spazi adeguati, come aule informatiche dedicate, fondamentali per un apprendimento efficace delle competenze digitali. Purtroppo, tali strutture sono spesso assenti o insufficienti in molte scuole italiane, limitando così le possibilità di formazione pratica e l'accesso a un'educazione digitale di qualità. Tuttavia, solo investendo in infrastrutture scolastiche adeguate sarà possibile creare un ambiente formativo che favorisca

---

<sup>55</sup> Cfr. sul punto L. PALAZZANI, *Digital Divide* (voce), in A.C. AMATO MANGIAMELI, G. SARACENI (a cura di), *Cento e una voce di informatica giuridica*, Giappichelli, 2023, 161; E. D'ORLANDO, *Profili costituzionali dell'Amministrazione digitale*, in *Diritto dell'Informazione e dell'Informatica*, 2, 2011, 220; E. DE MARCO, *Introduzione alla egualianza digitale*, in *Federalismi.it*, 12, 2008, 4 s.

<sup>56</sup> Il livello di competenze digitali in Italia è tra i più bassi d'Europa: cfr. Openpolis, *La sfida dell'alfabetizzazione digitale per contrastare le diseguaglianze*, 28 gennaio 2025.

lo sviluppo delle competenze digitali indispensabili per affrontare le sfide del presente e del futuro.

### **Considerazioni conclusive: la cybersecurity “metainteresse” nella tutela dei diritti fondamentali.**

È oggi attraverso il digitale che si assicura la fornitura di servizi essenziali per la collettività e che si consente alle aziende l'esercizio di gran parte delle attività imprenditoriali. La sicurezza delle reti è, quindi, il presupposto per l'erogazione di attività pubbliche e private su cui si basa non solo la nostra economia (come trapela spesso dalle normative europee di armonizzazione delle previsioni nazionali), ma anche il nostro Stato sociale.

Così come la sicurezza è stata già vista in passato come valore super-primario<sup>57</sup>, la cybersicurezza costituisce un “metainteresse” rispetto all'esercizio di taluni diritti fondamentali da parte dei singoli, che saranno tanto più restii al loro esercizio, quanto più è debole la loro *fiducia* nella sicurezza delle informazioni che li riguardano. In proposito, si deve riflettere sul fatto che la rinuncia all'accesso a determinate piattaforme equivale alla rinuncia all'accesso ai diritti e ai servizi.

Si può poi discutere sulla configurabilità di un “diritto fondamentale alla cybersicurezza”<sup>58</sup>, cioè di una qualificazione della cybersicurezza come di un “diritto al diritto di essere protetti”, proprio in quanto la disponibilità e integrità dei sistemi digitali sono condizioni preminenti per poter esercitare i propri diritti online.

Senza giungere necessariamente a tali conclusioni, è indubbio tuttavia che la cybersicurezza si configura come un vero e proprio “metainteresse”, ossia un interesse che sostiene l'attuazione di altri interessi: un prerequisito strutturale per il godimento dei diritti fondamentali. Senza sicurezza digitale, si minano libertà civili, accesso ai servizi, *privacy*, partecipazione democratica e fiducia nelle Istituzioni. Si può allora ben intuire quanto sia imprescindibile una preparazione sulle tematiche connesse alla garanzia dei sistemi di informazione e di rete, ma anche di base, rivolta a chi quotidianamente tratta dati, personali e non, nel contesto digitale.

La digitalizzazione, infatti, non è solo una questione tecnologica, ma incide profondamente sui diritti e sulle libertà fondamentali dei cittadini. Come è stato osservato in dottrina, infatti, «nella società digitale, [...] accade che ciascun diritto fondamentale finisca per assumere una dimensione bifronte, mostrando accanto al volto tradizionale, il suo doppio digitale, così che si possa parlare di identità personale e di identità digitale, di sicurezza e di

---

<sup>57</sup> G. CERRINA FERONI, G. MORBIDELLI, *La sicurezza: un valore superprimario*, in *Percorsi costituzionali*, 2008, 31 ss.

<sup>58</sup> Ritiene che non sia possibile parlare di un “diritto fondamentale alla cybersicurezza” E. LONGO, *Il diritto costituzionale e la cybersicurezza. Analisi di un volto nuovo del potere*, in *Rassegna parlamentare*, n. 2, 2024, 313 s. Contra P.G. CHIARA, *Towards a Right to Cybersecurity in EU Law? The Challenges Ahead*, in *Computer Law & Security Review*, 2023, online, 1 s.; G. CERRINA FERONI, G. MORBIDELLI, *La sicurezza: un valore superprimario*, cit., 31 s.

sicurezza informatica o cybersicurezza, di uguaglianza e di disuguaglianza digitale o *digital divide*, di doveri costituzionali di solidarietà e di solidarietà digitale e via dicendo»<sup>59</sup>.

Risulta dunque evidente che la cybersicurezza non può essere considerata un ambito esclusivamente tecnico o riservato agli operatori del settore informatico, ma si configura come una questione trasversale, che coinvolge direttamente la sfera dei diritti fondamentali e della cittadinanza digitale. In tale prospettiva, l’alfabetizzazione digitale rappresenta non solo uno strumento abilitante per la partecipazione consapevole alla vita digitale, ma anche un fattore determinante per la prevenzione e la mitigazione delle minacce informatiche. Una *cybersecurity* efficace non può prescindere, quindi, da una società digitalmente consapevole. Investire in programmi di alfabetizzazione digitale rappresenta oggi una condizione necessaria per garantire la resilienza del sistema digitale e la tutela sostanziale dei diritti nell’ambiente cibernetico.

È quindi urgente un ripensamento normativo che, integrando le misure di sicurezza tecnica con investimenti concreti in formazione e alfabetizzazione digitale, rafforzi la cultura della sicurezza informatica a tutti i livelli della società. Un’adeguata formazione non solo favorisce l’accesso e l’utilizzo consapevole degli strumenti digitali, ma riveste anche un ruolo cruciale nel rafforzare la consapevolezza riguardo ai rischi connessi alla condivisione di informazioni *online*, come la tutela della *privacy* e la sicurezza informatica. Solo attraverso l’educazione digitale sarà possibile formare cittadini preparati, capaci di navigare in modo sicuro e responsabile nell’ambiente digitale, contribuendo così a una società più inclusiva e resiliente.

## Bibliografia

AMORETTI F., GARGIULO E., *Dall’appartenenza materiale all’appartenenza virtuale. La cittadinanza elettronica fra processi di costituzionalizzazione della rete e dinamiche di esclusione*, in *Pol. dir.*, 3, 2010.

BARLOW J.P., *Dichiarazione d’indipendenza del Cyberspazio*, in *Duke Law & Technology Review*, 5-7, 2019.

BAVETTA F., *Direttiva NIS 2: verso un innalzamento dei livelli di cybersicurezza a livello europeo*, in *Media Laws*, 3, 2023.

BUFFA M., *La Direttiva NIS II Cybersecurity in Europa: tra innovazione, formazione e diritto vivente*, in *Democrazie e Diritti Sociali*, 1, 2023.

CALZOLAIO S., *Autorità indipendenti e di governo nella società digitale*, in F. PIZZETTI, S. CALZOLAIO, A. IANNUZZI, E. LONGO, M. OROFINO (a cura di), *La regolazione europea della società digitale*, cit.

CARAMASCHI O., *La cybersicurezza nazionale ai tempi della guerra (cibernetica): il ruolo degli organi parlamentari*, in *Osservatorio AIC*, 4, 2022.

<sup>59</sup> Così A. IANNUZZI, F. LAVIOLA, *I diritti fondamentali nella transizione digitale fra libertà e uguaglianza*, in *Dir. cost.*, 1, 2023, 12.

CARUSO C. *I custodi di silicio. Protezione della democrazia e libertà di espressione nell'era dei social network*, in AA.VV., *Liber Amicorum per Pasquale Costanzo*, Vol. I, in *Consulta Online*, 2020.

CASAROSA F., *L'armonizzazione degli obblighi di notifica: il DDL Cybersicurezza verso la NIS 2*, in *Rivista italiana di informatica e diritto*, 1, 2024.

CERRINA FERONI G., MORBIDELLI G., *La sicurezza: un valore superprimario*, in *Percorsi costituzionali*, 2008.

CHIARA P.G., BRIGHI R., *La dimensione della “resilienza” nel diritto Ue della cybersicurezza*, in *Ragion pratica*, 2, 2024.

CHIARA P.G., *Towards a Right to Cybersecurity in EU Law? The Challenges Ahead*, in *Computer Law & Security Review*, 2023, online.

DE MARCO E., *Introduzione alla egualanza digitale*, in *Federalismi.it*, 12, 2008.

FINOCCHIARO G., *Sovranità digitale*, in *Dir. Pubbl.*, 3, 2022.

GIUPPONI T. F., *Il governo nazionale della cybersicurezza*, in *Quad. Cost.*, 2, 2024.

IANNUZZI A., *Considerazioni sul disegno di legge «Disposizioni in materia di rafforzamento della cybersicurezza nazionale e di reati» (AC 1717)*, in *Rivista italiana di informatica e diritto*, 1, 2024.

IANNUZZI A., LAVIOLA F., *I diritti fondamentali nella transizione digitale fra libertà e uguaglianza*, in *Dir. cost.*, 1, 2023.

LONGO E., *Audizione informale per il disegno di legge in materia di «Disposizioni in materia di rafforzamento della cybersicurezza nazionale e di reati informatici» (AC 1717)*, in *Rivista italiana di informatica e diritto*, 1, 2024.

LONGO E., *Il diritto costituzionale e la cybersicurezza. Analisi di un volto nuovo del potere*, in *Rassegna parlamentare*, n. 2, 2024.

LONGO E., *La disciplina della cybersicurezza nell’Unione europea e in Italia*, in F. PIZZETTI, S. CALZOLAI, A. IANNUZZI, E. LONGO, M. OROFINO, *La regolazione europea della società digitale*, Giappichelli, Torino, 2024.

LOTTA C., *Governance della Rete, accesso a Internet e cybersicurezza*, Editoriale Scientifica, Napoli, 2024.

MENSI M., *La sicurezza cibernetica*, in M. MENSI, P. FALLETTA (a cura di), *Il diritto nel web*, Cedam, 2018.

MORONI L., *La governance della cybersicurezza a livello interno ed europeo: un quadro intricato*, in *Federalismi.it*, 14, 2024.

NANNIPIERI L., *Costituzione e nuove tecnologie: profili costituzionali dell’accesso a Internet*, in *Rivista dell’Associazione “Gruppo di Pisa”*, Secondo seminario annuale del “Gruppo di Pisa” con i dottorandi delle discipline giuridiche su “Lo studio delle fonti del diritto e dei diritti fondamentali in alcune ricerche dottorali”, Università degli Studi Roma Tre, 20 settembre 2013, 3.

PALAZZANI L., *Digital Divide* (voce), in A.C. AMATO MANGIAMELI, G. SARACENI (a cura di), *Cento e una voce di informatica giuridica*, Giappichelli, 2023, 161; E. D'ORLANDO, *Profili costituzionali dell'Amministrazione digitale*, in *Diritto dell'Informazione e dell'Informatica*, 2, 2011.

PARONA L., *L'istituzione dell'Agenzia per la cybersicurezza nazionale*, in *Giornale di diritto amministrativo*, 6, 2021.

PIETRANGELO M., *Per un modello nazionale di cybersicurezza cooperativa e resilienza collaborativa*, in *Rivista italiana di informatica e diritto*, 1, 2024.

POLETTI S., *La sicurezza cibernetica nazionale ed europea, alla luce della creazione del perimetro di sicurezza nazionale cibernetica*, in *Media Laws*, 2, 2023.

PREVITI L., *Pubblici poteri e cybersicurezza: il lungo cammino verso un approccio collaborativo alla gestione del rischio informatico*, in *Federalismi.it*, 25, 2022.

VENANZONI A., *L'ordine costituzionale della cybersecurity*, in *Forum Quad. Cost.*, 4, 2024.

## Allucinazioni dell'intelligenza artificiale: sintomi e cura – di *Giovanna Pistorio*, Professoressa associata in Diritto costituzionale e pubblico, presso il Dipartimento di Giurisprudenza dell'Università degli Studi Roma Tre

È il 14 marzo 2025 quando il Tribunale di Firenze, nel corso di un processo su plagio e riproduzione di marchio industriale, si pronuncia in tema di responsabilità aggravata per lite temeraria. La richiesta di condanna ex art. 96 c.p.c. è stata formulata a causa «dell'indicazione, in sede di comparsa di costituzione, di sentenze inesistenti, ovvero il cui contenuto reale non corrisponde a quello riportato»<sup>60</sup>.

Nulla di nuovo se non fosse che l'indicazione di tali decisioni non è il risultato della ricerca effettuata da un avvocato inesperto (o forse sì, *sic!*), ma è il frutto di allucinazioni di un sistema di intelligenza artificiale. Sulla base di una vaga somiglianza con quelle umane, le allucinazioni dell'intelligenza artificiale consistono in risposte contenenti informazioni incoerenti, fuorvianti o del tutto inesistenti.

È Chat GPT, nel caso di specie, a commettere l'errore. Si tratta della nota piattaforma *Generative Pre-trained Transformer* - come indica la sigla nel nome - che, sviluppata dalla società statunitense OpenAI nel 2020 e ormai a tutti nota, rappresenta la forma tecnologica più avanzata per la conversazione con una macchina.

Interrogata, dunque, nel corso di una ricerca giurisprudenziale, Chat GPT “inventa” alcuni precedenti che, non verificati poi in sede di redazione dell’atto giudiziario, entrano nel processo, rischiando di generare una visione distorta della realtà.

Nonostante il «disvalore relativo all’omessa verifica dell’effettiva esistenza delle sentenze risultanti dall’interrogazione dell’IA», il Tribunale di Firenze esclude l’applicabilità dell’art. 96 c.p.c. Non è dimostrata la mala fede della parte soccombente. L’utilizzo dell’intelligenza artificiale, ai fini della ricerca, è compiuto da una collaboratrice di studio, all’insaputa del patrocinatore. Non v’è prova dei danni subiti a causa dei riferimenti giurisprudenziali inesistenti, né del nesso di causalità tra condotta illecita e pregiudizio arrecato. Le sentenze errate vengono riportate nelle memorie difensive *ad abundantiam* rispetto alla strategia processuale già nota, sin dal primo grado di giudizio.

Nessuna lite temeraria, dunque, per assenza dei presupposti.

A prescindere dalla condivisibilità o meno dell’orientamento espresso dal Tribunale di Firenze, i profili di interesse che il caso suscita sono molti.

A cominciare da una prospettiva di ampio respiro, si pensi ai rischi che l’utilizzo delle nuove tecnologie implica<sup>61</sup>. Se è vero, infatti, che le meraviglie dell’intelligenza artificiale dischiudono nuovi scenari, fatti di importanti e significative opportunità – dalla globalizzazione delle relazioni all’accrescimento delle conoscenze – è altresì acclarato che esse non sono esenti da pericoli, sia in

---

<sup>60</sup> Tribunale di Firenze, sezione imprese, ordinanza pronunciata nel procedimento per reclamo iscritto al n. 11053/2024 R.G., 14 marzo 2025.

<sup>61</sup> È già dalla metà degli anni Novanta che S. RODOTÀ, *Tecnologie e diritti*, Bologna, 1995, si interrogava sulle possibili ripercussioni sulla sfera privata dell'uomo derivanti dalle nuove tecnologie. Poi, la dottrina è amplissima sul tema. Cfr., *ex plurumis*, U. BECK, *Conditio humana. Il rischio nell'età globale*, 2011.

termini di garanzia dei diritti individuali, dalla tutela della dignità alla privacy, sia rispetto a esigenze di interesse generale, ordine pubblico, amministrazione della giustizia, sicurezza<sup>62</sup>.

Con riguardo, in particolare, al caso di specie, poi, si pensi all'importanza della questione in termini di responsabilità professionale dell'avvocato. D'altra parte, se un legale, più o meno consapevolmente o addirittura dolosamente, manipola il contenuto di una sentenza, di un parere, di una fonte normativa, giungendo a una ricostruzione del tutto fuorviante rispetto all'originale, ne risponde personalmente per violazione dei doveri di lealtà e correttezza, oltretché, qualora ne ricorrano gli estremi, per falso o per altro reato.

In ogni caso, che l'errore sia parte della storia dell'umanità è un dato di fatto. Anzi, potremmo dire che l'errore nasce con l'uomo quando, acquisita la capacità di scelta, assume con essa il rischio di sbagliare. Come evocato nel libro della Genesi, Eva, tentata dal serpente, disobeisce mangiando il frutto dell'albero proibito. È un atto di disobbedienza, dunque, che porta l'umanità alla scelta e perciò anche alla possibilità di commettere errori.

Poi, nel tempo e nello spazio, gli errori si moltiplicano, evolvono, si sviluppano e diventano persino oggetto di studio e dibattito.

Per Platone l'errore è inevitabile e non necessariamente negativo, se prodromico a un processo di elevazione verso la verità. Per Aristotele, invece, l'errore è l'antitesi della verità, frutto di una distorta percezione della realtà e perciò solo da condannare, qualora non lo si riesca ad evitare.

Evitare gli errori è stato, in effetti, uno degli obiettivi cui mirava l'uomo quando ha iniziato a delegare a congegni meccanici determinate operazioni tipiche della mente. Dall'abaco alla calcolatrice per effettuare calcoli aritmetici, dagli algoritmi alle forme di intelligenza artificiale per compiere operazioni più complesse, in maniera rapida, efficiente, oggettiva e neutrale, come solo le macchine sono in grado di eseguire<sup>63</sup>.

Tutto o quasi sembra crollare, come un castello di sabbia, quando proprio quelle macchine che chiamiamo "intelligenti" cominciano a commettere errori, generando informazioni fuorvianti, capaci di incidere su calcoli statistici e analisi di dati. Nascono così le c.d. allucinazioni delle forme di intelligenza artificiale. Tali allucinazioni si risolvono in errori "innocui", come quando una chatbot consiglia una mostra a Parigi a chi chiede come trascorrere la serata a Campobasso, o in errori ben più gravi e significativi, se perpetrati in attività di consulenza medica, scientifica, finanziaria, giuridica.

Recenti studi, peraltro, dimostrano che l'incremento delle forme di allucinazioni risulti direttamente proporzionale allo sviluppo dei modelli delle macchine intelligenti. OpenAi ha pubblicato i dati di alcuni test interni, rivelando che se, all'inizio, la generazione di informazioni false si attestava intorno al 15%, successivamente è aumentata al 33% per arrivare, da ultimo, nei modelli più sofisticati, a toccare il

---

<sup>62</sup> Sulla possibilità che minacce ai diritti fondamentali e alla democraticità degli ordinamenti possano cumularsi, incrementando «il proprio potenziale offensivo», cfr. G. FONTANA, *Sfera pubblica digitale e democrazia dell'Unione europea. Prime considerazioni intorno alla dichiarazione europea sui diritti e i principi digitali*, in R. Torino, S. Zorzetto (a cura di), *La trasformazione digitale in Europa. Diritti e principi*, Torino, 2023, 41.

<sup>63</sup> Per una efficace decostruzione del discorso intorno «alla neutralità della macchina e al suo supposto agire in modo scevro da quei pregiudizi ed errori che inficiano le decisioni umane», si veda M. AIROLDI, D. GAMBETTA, *Sul mito della neutralità algoritmica*, in *The Lab's Quarterly*, 2018, 25 ss.

48%. La conseguenza, inevitabile quanto paradossale, è che al crescere della potenza si riduce l'affidabilità nei confronti delle macchine<sup>64</sup>.

È da tale contesto di indubbio disordine e preoccupazione che muovono le seguenti riflessioni: perché le macchine sbagliano e, soprattutto, a quali costi?

### Errori e discriminazioni delle macchine intelligenti

Per comprendere il motivo per cui le macchine commettono degli errori, occorre fare un passo indietro ed esaminarne il funzionamento.

L'intelligenza artificiale comprende un insieme eterogeneo e variegato di tecnologie in grado di acquisire, analizzare, pianificare, ovvero svolgere alcune delle attività tipicamente umane. Per compiere tali operazioni, la macchina si avvale di algoritmi. Gli algoritmi sono infatti quelle "istruzioni" che consentono di risolvere un problema tramite la rielaborazione di dati immessi nel sistema. Se poi, come avviene nel caso del *Machine Learning*, l'intelligenza artificiale viene programmata per imparare da un set di dati, la macchina, proprio a partire da tali dati, non solo apprende, ma inizia a generare dei contenuti in maniera autonoma. Ora, più è ampia la mole dei dati messa a disposizione, maggiore sarà la capacità di apprendimento della macchina e, conseguenzialmente, più precisa sarà la soluzione fornita dall'algoritmo al quesito<sup>65</sup>.

Tutto impeccabile, se non fosse per due zone d'ombra.

La prima è imputabile alle macchine e attiene alla possibile opacità della decisione algoritmica. La fase finale della sequenza di operazioni, rappresentata dall'elaborazione del risultato, sulla base dell'analisi dei dati, pur essendo «decisiva» è la «meno governabile»<sup>66</sup>. È qui che risiede quella che viene efficacemente definita la *black box*, posto che il meccanismo di correlazioni resta spesso indecifrabile anche per gli stessi programmati<sup>67</sup>.

Sono proprio i programmati, peraltro, che, se da un lato cercano di comprendere il funzionamento di tali passaggi, ovvero l'essenza dell'apprendimento automatico, dall'altro sono però i responsabili, sia pur in parte, della seconda zona d'ombra. La decisione algoritmica dipende infatti dalla progettazione dell'algoritmo e dalla configurazione dei set di dati. In entrambe queste fasi, il ruolo dell'uomo nell'interazione con le macchine è determinante. In sede di programmazione, è possibile ordinare al sistema la direzione, escludendo o includendo alcuni elementi o suggerendo la strategia da seguire. In sede di configurazione, poi, se i dati inseriti sono incoerenti, inesatti, obsoleti, la decisione assunta sarà inaffidabile, perché imprecisa, se non errata.

Vero è dunque che, a valle, il risultato finale è prodotto dalla macchina, ma è altresì indubbio che, guardando a monte, l'eventuale errore è imputabile all'addestramento con dati distorti o, comunque, non rappresentativi della realtà.

---

<sup>64</sup> Così, M. CARMIGNANI, *Le allucinazioni dell'IA aumentano: ecco perché*, in *Agendadigitale*, 9 giugno 2025.

<sup>65</sup> In tal senso, cfr. P. ZUDDAS, *Intelligenza artificiale e discriminazioni*, in [www.giuricost.org](http://www.giuricost.org), 2020, 2.

<sup>66</sup> Così, P. ZUDDAS, *Intelligenza artificiale e discriminazioni*, cit., 11.

<sup>67</sup> Più approfonditamente, cfr. D.U. GALETTA-J.G. CORVALÁN, *Intelligenza Artificiale per una Pubblica Amministrazione 4.0? Potenzialità, rischi e sfide della rivoluzione tecnologica in atto*, in *Federalismi.it*, 2019, 15 s.

Detto in altri termini, in tali casi, la responsabilità è dell'uomo.

Avalendosi di categorie proprie del diritto penale, potremmo differenziare il comportamento, in base all'elemento soggettivo: dolo o colpa.

Ricorre il dolo quando ci si avvale di algoritmi per influenzare potenziali utenti, inducendoli a determinati comportamenti, per fini commerciali o politici, nell'acquisto di un prodotto o in campagna elettorale, per esempio. È a tutti noto lo scandalo di Cambridge analytica: società di consulenza che raccoglieva e analizzava dati, ricavati per lo più da social network, da Facebook in particolare, per creare messaggi pubblicitari personalizzati e influenzare così gli elettori nelle loro scelte politiche<sup>68</sup>.

Diverso, invece, il caso in cui la distorsione cognitiva della macchina è frutto di un atteggiamento colposo dell'uomo: negligenza, imperizia, mancanza di controllo.

L'evenienza più frequente e pregiudizievole è quella dei c.d. *bias*, vale a dire distorsioni che portano poi a errori di calcolo, di valutazione, di scelta.

Facciamo qualche esempio.

Uno dei primi casi risale addirittura agli anni Ottanta, quando nel Regno Unito, in una Scuola di medicina, ci si avvalse di un sistema di intelligenza artificiale per la selezione del personale: il sistema, operando sulla base delle assunzioni effettuate nel passato, finì per penalizzare i candidati di sesso femminile e i candidati immigrati<sup>69</sup>.

Uno dei casi più noti, invece, si verifica negli Stati Uniti, intorno agli anni 2010, quando comincia ad essere utilizzato il software COMPASS (*Correctional offender management profiling for alternative sanctions*), come supporto al sistema giudiziario, essendo uno strumento per calcolare, sulla base di alcuni indici, il rischio di recidiva di eventuali condannati. Qual è il problema? Il sistema è stato programmato sulla base di informazioni e decisioni anche piuttosto datate che, inevitabilmente, incorporavano indubbi pregiudizi. Ecco allora che, nel 2013, Loomis, condannato di sesso maschile, afroamericano, appartenente a un ceto sociale basso, ottiene il più elevato rischio di recidiva<sup>70</sup>. Evidente la discriminazione, intersezionale nel caso di specie, ovvero derivante dall'interazione di più discriminazioni – sesso, razza, estrazione sociale – quale effetto della configurazione dei dati<sup>71</sup>. Una discriminazione *by design*, potremmo dire.

Analogo pregiudizio pervade, spesso, i sistemi sanitari. Risale al 2022, la ricerca condotta nell'Università di Boston che ha dimostrato quanti e quali pregiudizi, nascosti nei dispositivi clinici, sono il frutto di progettazioni inique, per etnia, genere, età<sup>72</sup>. Nella diagnosi e nella cura di malattie cardiovascolari, per esempio, è ormai acclarato che molte delle distorsioni cognitive che hanno generato discriminazioni in base al sesso sono imputabili alla progettazione degli algoritmi, non calibrata in modo tale da valutare adeguatamente le differenze di rischio tra uomo e donna e in relazione

---

<sup>68</sup> Interessanti, quanto problematiche, le ripercussioni in termini di privacy, sicurezza e disinformazione, esaminate, con riguardo al caso in esame, da E. ASSANTE, *Cosa ci può insegnare il caso Cambridge Analytica*, in *Federalismi*, 2019.

<sup>69</sup> Più approfonditamente, cfr. S. LOWRY, G. MACPHERSON, *A blot on the profession*, in *296 British Medical Journal*, 1988, 657 ss.

<sup>70</sup> Supreme Court del Wisconsin nel caso *State v. Loomis*, 881 N.W.2d 749, 2016.

<sup>71</sup> Sul caso, *ex multis*, A. SIMONCINI, S. SUWEIS, *Il cambio di paradigma nell'intelligenza artificiale e il suo impatto sul diritto costituzionale*, in *Rivista di filosofia del diritto*, 1/2019, 102.

<sup>72</sup> In tal senso, cfr. P. COZZI, *Nuove riflessioni critiche sul rischio di discriminazione algoritmica in ambito sanitario*, in *Tech4future*, 2025.

all'età<sup>73</sup>. Negli Stati Uniti ci si è affidati a una forma di intelligenza artificiale per prevedere quali pazienti avrebbero necessitato di ricevere determinate cure mediche. In tal caso però, posto che i dati venivano raccolti basandosi sulla spesa sanitaria e acclarato che i pazienti neri sostenevano spese sanitarie inferiori rispetto ai bianchi, nelle stesse condizioni di salute, si è generata una discriminazione etnica, per effetto della quale venivano selezionati, per usufruire di dette cure, per lo più pazienti bianchi.

Non è esente da rischi di questo tipo il sistema bancario. Moltissime le discriminazioni perpetrare nei confronti di gruppi a basso reddito, da algoritmi finanziari che, inadeguatamente progettati, hanno processato le richieste di mutuo<sup>74</sup>.

In tali occasioni, dunque, la decisione algoritmica determina pregiudizi digitali «derivati», vale a dire frutto di scelte discriminatorie imputabili all'intervento umano, in sede di progettazione o di configurazione dei dati<sup>75</sup>. D'altra parte, se i dati sono «soggettivamente inquinati» da pregiudizi, tali pregiudizi sono destinati a cristallizzarsi<sup>76</sup>. Incorporata la distorsione, la macchina non può che perpetuare<sup>77</sup>, se non aggravare le discriminazioni<sup>78</sup>.

## I “costi” delle allucinazioni

La possibile propagazione, nel tempo e nello spazio, delle discriminazioni perpetuate dall'intelligenza artificiale, ma derivate dall'intervento umano, induce a riflettere sugli effetti, in termini di “costi”, di tali pregiudizi. Le discriminazioni, spesso intersezionali, derivanti da operazioni algoritmiche appaiono infatti molto più gravi rispetto a quelle realizzate senza l'uso dell'intelligenza artificiale.

Almeno tre, le ragioni.

In primo luogo, tali decisioni hanno un'applicazione quantitativamente significativa, una capacità diffusiva immediata e capillare. Si tratta, d'altra parte, di una delle inedite opportunità per l'uomo e per

---

<sup>73</sup> In tal senso, cfr. *AI e Bias: i pregiudizi trasmessi dall'umano alla macchina*, in [www.jnjmedicalcloud.it](http://www.jnjmedicalcloud.it), 2024.

<sup>74</sup> Più approfonditamente, cfr. M. MARTORANA, *Banche: i danni dell'intelligenza artificiale: discriminazioni ed errori*, in [www.agendadigitale.it](http://www.agendadigitale.it), 2021.

<sup>75</sup> Per tale orientamento, cfr. P. ZUDDAS, *Intelligenza artificiale e discriminazioni*, cit., 5.

<sup>76</sup> Così, A. VENANZONI, *La valle del perturbante: il costituzionalismo alla prova delle intelligenze artificiali e della robotica*, in *Politica del diritto*, 2019, 237 ss.

<sup>77</sup> Più approfonditamente, sulla configurazione degli algoritmi, quali prodotti storici che riflettono scelte della società, cfr. N. SEAVER, *Algorithms as culture: Some tactics for the ethnography of algorithmic systems*, 2017, in *Big Data & Society*, 2 ss.; T. STRIPHAS, *Algorithmic culture*, in *European Journal of Cultural Studies*, 2015, 395 ss.

<sup>78</sup> Interessante ed emblematico in tal senso l'esempio riportato da G. RESTA, *Governare l'innovazione tecnologica: decisioni algoritmiche, diritti digitali e principio di uguaglianza*, in *Politica del diritto*, 2019, 217, secondo il quale «se nella programmazione di un algoritmo utilizzato per decisioni relative all'assunzione di personale si concretizza la nozione di «buon» dipendente avendo riguardo – tra gli altri – al criterio della puntualità nel recarsi sul luogo di lavoro, ciò può finire per penalizzare sistematicamente tutti coloro che vivono in periferia ed impiegano di conseguenza maggior tempo per raggiungere ogni giorno la sede dell'impresa. E poiché in determinati contesti la circostanza dell'abitare in periferia è una variabile strettamente correlata alle origini etniche e alle condizioni sociali più disagiate, un siffatto criterio formalmente “neutro” potrebbe finire per riflettersi negativamente a danno di categorie già svantaggiate».

la società intera, derivanti dallo sviluppo delle nuove tecnologie, dal progresso scientifico, tecnologico e informatico<sup>79</sup>.

In secondo luogo, è proprio tale progresso che spiega la maggiore gravità dei pregiudizi arrecati dagli algoritmi. Si tratta infatti di un progresso che affascina e conquista l'uomo per i risultati da esso prodotti. Ne consegue uno slancio, spesso eccessivamente fideistico, verso le meraviglie dell'intelligenza artificiale. L'uomo crede, confida nelle macchine, *rectius* si affida ai sistemi di intelligenza artificiale, con una propensione spiccatamente maggiore rispetto all'atteggiamento di apertura nei confronti dei suoi simili, finanche professionisti o esperti. L'algoritmo è spesso visto come «uno stabilizzatore di fiducia», garanzia di efficienza e di neutralità<sup>80</sup>.

Basti pensare alla capacità di persuasione, non tanto e non solo dell'informazione in rete, ma della disinformazione. Le *fake news* sono sempre esistite, ma l'irrompere della rete sulla scena ha inciso profondamente sulle modalità di diffusione delle stesse, incrementandone da un punto di vista quantitativo e qualitativo il raggio di espansione e la capacità di condizionamento<sup>81</sup>. Tutto, o quasi, in rete sembra vero, o quantomeno verosimile e quindi credibile. D'altra parte, è proprio attraverso la rete, quell'ambiente artificiale, parallelo alla vita reale, senza confini di spazio e di tempo, che l'internet delle cose è ormai penetrato nella vita quotidiana, caratterizzandone le abitudini, alterandola, incidendo sulla stessa in modo significativo<sup>82</sup>. Oggi assistiamo a quella che è stata efficacemente definita la quarta rivoluzione, ovvero quella in cui le nuove tecnologie forgiano la realtà fisica e intellettuale dell'uomo, incidendo sul modo in cui questi, come organismo informazionale, si relaziona con gli altri e con se stesso in un ambiente *onlife* del tutto nuovo e dall'evoluzione costante<sup>83</sup>. La regolazione della vita sociale è ormai incardinata entro i confini determinati dall'universo digitale in cui sequenze di algoritmi irrompono silenziosamente nella vita di ciascun individuo<sup>84</sup>, indagando nei settori più disparati: dallo stato di salute, agli interessi, alle scelte economiche, alla potenzialità criminale, all'efficienza lavorativa<sup>85</sup>. Le macchine intelligenti consentono di archiviare, sistemare, ordinare una quantità enorme di dati e di informazioni, realizzando una precisa e analitica profilazione del singolo e delle collettività<sup>86</sup>. La profilazione consente di delineare l'identità virtuale di una persona, individuandone così abitudini, idee, volontà<sup>87</sup>.

Gli effetti che ne scaturiscono consentono di comprendere la terza ragione per cui le discriminazioni derivanti dai sistemi di intelligenza artificiale sono più gravi e pregiudizievoli. Gli algoritmi, sfruttando

<sup>79</sup> Più approfonditamente, cfr. L. PALAZZANI, *Tecnologie dell'informazione e intelligenza artificiale*, Roma, 2020, 7 ss.

<sup>80</sup> In tal senso, cfr. T. GILLESPIE, *The relevance of algorithms*, in T. Gillespie, P. Boczkowski, K. Foot (a cura di) *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge, 2014, 179.

<sup>81</sup> In tal senso, cfr. C. CARUSO, *Il tempo delle istituzioni di libertà. Piattaforme digitali, disinformazione e discorso pubblico europeo*, in *Quaderni costituzionali*, 2023, 545 ss.

<sup>82</sup> Sul tema, cfr. N. LETTIERI, *Antigone e gli algoritmi. Appunti per un approccio giusfilosofico*, Modena, 2020, 13 ss.

<sup>83</sup> In tal senso, cfr. L. FLORIDI, *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, Milano, 2017, 99 ss.

<sup>84</sup> Secondo N. BOBBIO, *L'età dei diritti*, Tirino, 1990, 263, «oggi le minacce alla vita, alla libertà e alla sicurezza possono venire dal potere sempre più grande che le conquiste della scienza e delle applicazioni che ne derivano danno a chi è in condizione di usarne».

<sup>85</sup> Più approfonditamente, cfr. A. ODDENINO, *Individual Rights, Cybersecurity and Access to Knowledge on the Internet: The Revision of ITU International Telecommunication Regulations*, in *Diritti umani e diritto internazionale*, 2013, 532 ss. T.Z. ZARSKY, *An analytic challenge: discrimination theory in the age of predictive analytics*, in *A Journal of Law and Policy for the Information Society*, 2017, 12.

<sup>86</sup> È già dalla metà degli anni Novanta che S. RODOTÀ, *Tecnologie e diritti*, Bologna, 1995, si interrogava sulle possibili ripercussioni sulla sfera privata dell'uomo derivanti dalle modalità informatiche e tecnologiche di raccolta dei dati.

<sup>87</sup> In tal senso, cfr. M. HILDEBRANDT, *Defining profiling: a new type of knowledge?* in M. Hildebrandt-S. Gutwirth (a cura di), *Profiling the European citizen*, Springer, Berlin, 2008.

enormi volumi di dati, più o meno consapevolmente immessi in rete o nelle macchine dall'uomo, influenzano le persone, ne plasmano le decisioni nelle scelte di vita. Alle volte, non è il cittadino a cercare l'informazione, ma è la notizia a cercare il cittadino, nella veste di utente, acquirente, consumatore, producendo un effetto di "isolamento" nel mondo informativo, di riduzione dello spazio di conoscenza, di polarizzazione delle opinioni a discapito di quella che dovrebbe essere la pluralità delle espressioni, quale cartina di tornasole di un sistema democratico<sup>88</sup>. La profilazione, poi, se da un lato può essere utile per attività di controllo, poste in essere dai pubblici poteri, per motivi di difesa, sicurezza e prevenzione, dall'altro può invece divenire lo strumento per la commissione di reati di vario genere, tra i quali furti di identità, manipolazione di dati, frodi informatiche di vario tipo. È il c.d "rischio illecito", ovvero la possibilità che il trattamento dei dati personali non avvenga conformemente alla legge, provocando «danni fisici, materiali o immateriali»<sup>89</sup>.

### **Evitare l'algocrazia e “curare” le allucinazioni**

Il contesto è complicato e, per certi versi, pericoloso anche sul piano etico, prima ancora che giuridico. Dietro il rischio che l'«algocrazia» pone, ovvero il timore che gli algoritmi plasmando, se non sostituendo le decisioni umane, prendano il sopravvento, si sviluppa una questione di etica degli algoritmi, intesa quale responsabilità nella programmazione e nell'utilizzo delle forme di intelligenza artificiale<sup>90</sup>.

D'altra parte, una cosa è certa: tutto, o quasi, ruota attorno al rapporto tra uomo e macchina.

Posto che, in tale relazione, è fondamentale la modalità attraverso la quale l'uomo interagisce con i sistemi di intelligenza artificiale, il principio di non esclusività algoritmica deve costituire la base di tale rapporto e orientarne lo sviluppo. Sancito dal *General Data Protection Regulation* (GDPR, di cui al Regolamento UE n. 2016/679) e ribadito nell'*Artificial Intelligence Act* (AI Act, entrato in vigore il 1º agosto 2024), tale principio implica, da un lato, il diritto a non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, dall'altro, il diritto all'intervento umano. Siffatto diritto, in un'epoca efficacemente definita «Novacene», perché caratterizzata dalla coesistenza di uomini con dispositivi intelligenti<sup>91</sup>, assume carattere universale<sup>92</sup>, nel rapporto fra uomo e macchina, garantendo l'individuazione di attività che sono e devono rimanere «riserva di umanità»<sup>93</sup>.

Prima fra tutte, la supervisione, *ex ante* ed *ex post*.

---

<sup>88</sup> Più approfonditamente, E. PARISER, *Filter Bubble*, London, 2011, 15.

<sup>89</sup> Così, D. MULA, *Protezione dei dati personali e cybersecurity*, in *Cybersecurity Law. Disciplina italiana ed europea della sicurezza cibernetica anche alla luce delle norme tecniche*, a cura di A. Contaldo e D. Mula, Pisa, 2020, 155.

<sup>90</sup> Più approfonditamente, cfr. A. CELOTTO, *Come regolare gli algoritmi. Il difficile bilanciamento tra scienza, etica e diritto*, in *Analisi giuridica dell'Economia*, 2019, 48 s.

<sup>91</sup> Così, J. LOVELOCK, *Novacene. L'età dell'iperintelligenza*, Milano, 2020.

<sup>92</sup> In tal senso, cfr. A. IANNUZZI, *Le fonti del diritto dell'Unione europea per la disciplina della società digitale*, in AA.VV., *La regolazione europea della società digitale*, Torino, 2024, 26.

<sup>93</sup> In tal senso, G. GALLONE, *Riserva di umanità e funzioni amministrative. Indagini sui limiti all'automazione decisionale tra procedimento e processo*, Milano, 2023.

È a monte, infatti, che il controllo garantito dall'intervento umano, in sede di programmazione e di configurazione, per esempio, può evitare che il sistema di apprendimento incorpori pregiudizi e produca poi inevitabili discriminazioni<sup>94</sup>.

Discriminazioni che, comunque, potrebbero essere smascherate, a valle, tramite un'attenta e scrupolosa attività di supervisione umana, prodromica all'utilizzo dei risultati prodotti dalle macchine intelligenti. Solo tale attività, peraltro, è in grado di arginare gli effetti di eventuali *allucinazioni* imputabili alla macchina, nei margini di autonomia in cui, operando, può commettere errori.

Il caso deciso dal Tribunale di Firenze ne è un esempio emblematico. Il primo in Italia, a quanto costa, ma non in assoluto. Nel 2023, negli Stati Uniti, in un processo intentato per un risarcimento da lesioni personali, l'avvocato Steven A. Schwartz porta, a sostegno della domanda, diversi precedenti rivelatisi inesistenti e falsi, perché "inventati" da Chat GPT. Nessun dolo per il legale, posto che si dichiara ignaro del fenomeno delle allucinazioni e assolutamente privo di alcuna intenzione di ingannare il tribunale o la controparte. Tuttavia, accertata la colpa, per omesso controllo in merito alla veridicità delle informazioni reperite, l'avvocato viene sanzionato<sup>95</sup>.

Ecco allora che, per assicurare che, nel rapporto tra uomo e macchina, non siano gli algoritmi o i valori di coloro che li generano a "governare", ma venga assicurato «un adeguato bilanciamento tra la "funzionalità tecnologica" e la desiderabilità sociale degli scopi perseguiti», occorre investire su tre fronti<sup>96</sup>.

Incrementare la *cultura del digitale*, in termini di educazione e di alfabetizzazione, per un uso consapevole delle macchine intelligenti<sup>97</sup>. D'altra parte, quella riferita alla tecnologia è la più insidiosa delle sfide che il Ventunesimo secolo attende il costituzionalismo, a fronte dei dirompenti effetti che l'innovazione digitale produce sulla tenuta dell'assetto democratico<sup>98</sup>.

Potenziare la *cultura dell'uguaglianza*, in termini di equità, perché se è vero che tanto si è fatto, c'è ancora molta strada da percorrere: deve ancora cambiare il contesto sociale in cui si programma l'algoritmo perché molte discriminazioni delle forme di intelligenza artificiale sono lo specchio dei pregiudizi della società<sup>99</sup>. Come testimoniato da una ricerca condotta da Ryan Steed (Carnegie Mellon

---

<sup>94</sup> Più approfonditamente, cfr. K. CRAWFORD, *The Hidden Biases in Big Data*, in *Harvard Business Review*, 2013, 1 ss.

<sup>95</sup> Cfr. United States District Court, S.D. New York, *Mata v. Avianca Inc*, 22 giugno 2023.

In Italia, degno di nota – se non altro per l'esigenza di dover mettere per iscritto siffatta indicazione - il disegno di legge in materia di intelligenza artificiale approvato dal Senato lo scorso 20 marzo e attualmente in prima lettura alla Camera, nella parte in cui prevede, all'art. 13, che «l'utilizzo dei sistemi di intelligenza artificiale, nelle professioni intellettuali, è finalizzato al solo esercizio delle attività strumentali e di supporto all'attività professionale e con prevalenza del lavoro intellettuale oggetto della prestazione d'opera».

<sup>96</sup> Così, G. RESTA, *Governare l'innovazione tecnologica: decisioni algoritmiche, diritti digitali e principio di uguaglianza*, cit., 233.

<sup>97</sup> In tal senso, cfr. D. PEDRESCHI, F. GIANNOTTI, R. GUIDOTTI, A. MONREALE, L. PAPPALARDO, S. RUGGIERI, F. TURINI, *Open the Black Box: Data-Driven Explanation of Black Box Decision Systems*, in *ArXiv*, 2018, 26, secondo i quali è fondamentale investire in tecnologie e metodologie per spiegare l'attività degli algoritmi, salvaguardando autonomia e consapevolezza umana.

<sup>98</sup> Più approfonditamente, cfr. V. FROSINI, *Liberté égalité internet*, Napoli, 2019, 189 ss.

<sup>99</sup> Sul tema, cfr. M. D'AMICO, C. NARDOCCI, *Intelligenza artificiale e discriminazione di genere: rischi e possibili soluzioni*, in G. Cerrina Feroni, C. Fontana, E. C. Raffiotta (a cura di), *AI Anthology. Profili giuridici, economici e sociali dell'intelligenza artificiale*, Bologna, 2022, 251 ss.

University) e Aylin Caliskan (George Washington University), sempre più spesso l'«algoritmo diventa "portavoce" di pregiudizi e stereotipi esistenti»<sup>100</sup>.

Garantire il costante e diligente *intervento umano*, quale «prima concretizzazione dell'umanesimo digitale»<sup>101</sup>, propugnata dall'Unione europea. In tale contesto, assume rilevanza strategica la Dichiarazione europea sui diritti e i principi digitali per il decennio digitale, adottata da Parlamento, Consiglio e Commissione dell'Ue il 15 dicembre 2022. In piena armonia con precedenti scelte normative<sup>102</sup>, la Dichiarazione si prefigge, sin dal Preambolo, di «promuovere un modello europeo per la trasformazione digitale, che metta al centro le persone, (...) basato sui valori europei e sui diritti fondamentali dell'UE [e che] riaffermi i diritti umani universali e apporti benefici a tutte le persone, alle imprese e alle società nel suo complesso»<sup>103</sup>. La visione antropocentrica, quale filo conduttore della Dichiarazione, ponendosi peraltro in perfetta sintonia con l'impostazione valoriale della nostra Costituzione che pone l'uomo al centro della società, è e resta l'arma più efficace per evitare che la rivoluzione epocale in corso dia vita un «nuovo medioevo digitale»<sup>104</sup>.

## Bibliografia

*AI e Bias: i pregiudizi trasmessi dall'umano alla macchina*, in [www.jnjmedicalcloud.it](http://www.jnjmedicalcloud.it), 2024.

AIROLDI M., GAMBETTA D., *Sul mito della neutralità algoritmica*, in *The Lab's Quarterly*, 2018.

ASSANTE E., *Cosa ci può insegnare il caso Cambridge Analytica*, in *Federalismi*, 2019.

BALDRATI A., *Se l'algoritmo diventa lo specchio dei nostri stereotipi: uno studio*, in *Agendadigitale*, 26 febbraio 2021.

BECK U., *Conditio humana. Il rischio nell'età globale*, 2011.

BOBBIO N., *L'età dei diritti*, Torino, 1990.

CARMIGNANI M., *Le allucinazioni dell'IA aumentano: ecco perché*, in *Agendadigitale*, 9 giugno 2025.

CARUSO C., *Il tempo delle istituzioni di libertà. Piattaforme digitali, disinformazione e discorso pubblico europeo*, in *Quaderni costituzionali*, 2023.

CELESTE E., *Towards a European Declaration on Digital Rights and Principles: Guidelines for the Digital Decade*, in *dcubexitinstitute.eu*, 7 Febbraio 2022.

---

<sup>100</sup> Così, A. BALDRATI, *Se l'algoritmo diventa lo specchio dei nostri stereotipi: uno studio*, in *Agendadigitale*, 26 febbraio 2021.

<sup>101</sup> Così, A. IANNUZZI, *Le fonti del diritto dell'Unione europea per la disciplina della società digitale*, cit., 27.

<sup>102</sup> Con riferimento al *General Data Protection Regulation* (GDPR), al *Digital Service Act* (DSA), al *Digital Market Act* (DMA), all'*Artificial Intelligence Act* (AI Act), cfr., più approfonditamente, G. FONTANA, *Sfera pubblica digitale e democrazia dell'Unione europea. Prime considerazioni intorno alla dichiarazione europea sui diritti e i principi digitali*, cit., 66 ss.

<sup>103</sup> Sul punto, cfr. E. CELESTE, *Towards a European Declaration on Digital Rights and Principles: Guidelines for the Digital Decade*, in *dcubexitinstitute.eu*, 7 Febbraio 2022. L. CIANCI, *Dichiarazione europea sui diritti e i principi digitali: quid pluris?*, in *Dir. pubbl. comp. eur.*, 2/2022, 381 ss.

<sup>104</sup> Per tale orientamento, cfr. G. RESTA, *Governare l'innovazione tecnologica: decisioni algoritmiche, diritti digitali e principio di uguaglianza*, cit., 233.

CELOTTO A., *Come regolare gli algoritmi. Il difficile bilanciamento tra scienza, etica e diritto*, in *Analisi giuridica dell'Economia*, 2019.

CIANCI L., *Dichiarazione europea sui diritti e i principi digitali: quid pluris?*, in *Dir. pubbl. comp. eur.*, 2/2022.

COZZI P., *Nuove riflessioni critiche sul rischio di discriminazione algoritmica in ambito sanitario*, in *Tech4future*, 2025.

CRAWFORD K., *The Hidden Biases in Big Data*, in *Harvard Business Review*, 2013.

D'AMICO M., NARDOCCI C., *Intelligenza artificiale e discriminazione di genere: rischi e possibili soluzioni*, in G. Cerrina Feroni, C. Fontana, E. C. Raffiotta (a cura di), *AI Anthology. Profili giuridici, economici e sociali dell'intelligenza artificiale*, Bologna, 2022.

FLORIDI L., *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, Milano, 2017.

FONTANA G., *Sfera pubblica digitale e democrazia dell'Unione europea. Prime considerazioni intorno alla dichiarazione europea sui diritti e i principi digitali*, in R. Torino, S. Zorzetto (a cura di), *La trasformazione digitale in Europa. Diritti e principi*, Torino, 2023.

FROSINI V., *Liberté égalité internet*, Napoli, 2019.

GALETTA D. U. – CORVALÁN J.G., *Intelligenza Artificiale per una Pubblica Amministrazione 4.0? Potenzialità, rischi e sfide della rivoluzione tecnologica in atto*, in *Federalismi.it*, 2019.

GALLONE G., *Riserva di umanità e funzioni amministrative. Indagini sui limiti all'automazione decisionale tra procedimento e processo*, Milano, 2023.

GILLESPIE T., *The relevance of algorithms*, in T. Gillespie, P. Boczkowski, K. Foot (a cura di) *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge, 2014.

HILDEBRANDT M., *Defining profiling: a new type of knowledge?* in M. Hildebrandt-S. Gutwirth (a cura di), *Profiling the European citizen*, Springer, Berlin, 2008.

IANNUZZI A., *Le fonti del diritto dell'Unione europea per la disciplina della società digitale*, in AA.Vv., *La regolazione europea della società digitale*, Torino, 2024.

LETTIERI N., *Antigone e gli algoritmi. Appunti per un approccio giusfilosofico*, Modena, 2020.

LOVELOCK J., *Novacene. L'età dell'iperintelligenza*, Milano, 2020.

LOWRY S., MACPHERSON G., *A blot on the profession*, in 296 *British Medical Journal*, 1988.

MARTORANA M., *Banche: i danni dell'intelligenza artificiale: discriminazioni ed errori*, in [www.agendadigitale](http://www.agendadigitale.it), 2021.

MULA D., *Protezione dei dati personali e cybersecurity*, in *Cybersecurity Law. Disciplina italiana ed europea della sicurezza cibernetica anche alla luce delle norme tecniche*, a cura di A. Contaldo e D. Mula, Pisa, 2020.

ODDENINO A., *Individual Rights, Cybersecurity and Access to Knowledge on the Internet: The Revision of ITU International Telecommunication Regulations*, in *Diritti umani e diritto internazionale*, 2013.

PALAZZANI L., *Tecnologie dell'informazione e intelligenza artificiale*, Roma, 2020.

PARISER E., *Filter Bubble*, London, 2011.

PEDRESCHI D., GIANNOTTI F., GUIDOTTI R., MONREALE A., PAPPALARDO L., RUGGIERI S., TURINI F., *Open the Black Box: Data-Driven Explanation of Black Box Decision Systems*, in *ArXiv*, 2018.

RESTA G., *Governare l'innovazione tecnologica: decisioni algoritmiche, diritti digitali e principio di uguaglianza*, in *Politica del diritto*, 2019.

- RODOTÀ S., *Tecnologie e diritti*, Bologna, 1995.
- SEAVER N., *Algorithms as culture: Some tactics for the ethnography of algorithmic systems*, 2017, in *Big Data & Society*.
- SIMONCINI A., SUWEIS S., *Il cambio di paradigma nell'intelligenza artificiale e il suo impatto sul diritto costituzionale*, in *Rivista di filosofia del diritto*, 1/2019.
- STRIPHAS T., *Algorithmic culture*, in *European Journal of Cultural Studies*, 2015.
- VENANZONI A., *La valle del perturbante: il costituzionalismo alla prova delle intelligenze artificiali e della robotica*, in *Politica del diritto*, 2019.
- ZARSKY T.Z., *An analytic challenge: discrimination theory in the age of predictive analytics*, in *A Journal of Law and Policy for the Information Society*, 2017.
- ZUDDAS P., *Intelligenza artificiale e discriminazioni*, in [www.giuricost.org](http://www.giuricost.org), 2020.

## **Il GDPR nell'era di Big Data e IA. Riflessioni su alcune oscillazioni tra obsolescenza e lungimiranza - di *Teresa Baldazzi* Dottoranda in Law&Social Change presso il Dipartimento di Giurisprudenza dell'Università degli Studi Roma Tre. Esperta con orientamento in Diritto delle nuove tecnologie presso l'Agenzia per la cybersicurezza nazionale**

*Big Data e intelligenza artificiale accelerano l'obsolescenza normativa e mettono alla prova il GDPR. Tra definizioni instabili, re-identificazione e “scatole nere”, il diritto cerca nuovi punti d'appoggio: dalla privacy by design alle garanzie sulle decisioni automatizzate, fino al raccordo con l'AI Act*

Mentre l'umanità scivola in uno stato di «dipendenza sistematica e strutturale dalla datificazione»<sup>105</sup> e assiste alla (o, meglio, produce la) graduale trasformazione delle macchine da *strumenti ad agenti*<sup>106</sup>, il diritto fatica a tenere il passo. A bene vedere, ciò è, in parte, strutturale, se è vero che il diritto è la risposta che nasce, con Ulrich Beck, «dalla riflessione pubblica sull'orrore prodotto dalla vittoria della modernità»<sup>107</sup>. Per altra parte, tuttavia, sono l'incessante e repentina evoluzione tecnologica ad aver accelerato il fenomeno dell'obsolescenza normativa, sospingendo, al contempo, l'ordinamento verso forme di regolazione soft e di auto- e co-regolazione<sup>108</sup>.

Poiché sembra tuttora potersi affermare che la chiave di volta del sistema di tutele della persona nell'*infosfera*<sup>109</sup> sia il diritto alla riservatezza – in particolare sotto il profilo dinamico di diritto alla protezione dei dati personali e all'autodeterminazione informativa –, il contributo mira a indagare la “tenuta” del Regolamento (UE) 679/2016 (*General Data Protection Regulation*, GDPR), a fronte dell'emersione di nuovi fenomeni tecnologici, in particolare di *Big Data* e dell'*Intelligenza Artificiale* (IA).

Il problema di fondo è che il Regolamento in parola, pur non ignorando le profonde trasformazioni che già erano in atto e, anzi, ponendo norme, per certi aspetti, decisamente

---

<sup>105</sup> S. CALZOLAIO, *Vulnerabilità della società digitale e ordinamento costituzionale dei dati*, in *Rivista italiana di informatica e diritto*, n. 2/2023, 19.

<sup>106</sup> A. SIMONCINI, S. SUWEIS, *Il cambio di paradigma nell'intelligenza artificiale e il suo impatto sul diritto costituzionale*, in *Rivista di filosofia del diritto*, n. 1/2019, 87 ss.

<sup>107</sup> U. BECK, *The Metamorphosis of the World*, Cambridge, Polity Press, 2016, tr. it. di Marco Cupellaro, *La metamorfosi del mondo*, Bari, Editori Laterza, 2017, 41.

<sup>108</sup> Sul tema v. Atti del Convegno finale del Progetto PRIN 2017 *Self- and Co-regulation for Emerging Technologies: Towards a Technological Rule of Law* (SE.CO.R.E. TECH.), Firenze 8-9 febbraio 2024, in *Osservatorio sulle fonti*, Speciale n. 1/2024.

<sup>109</sup> Cioè, nelle parole di L. Floridi, quell’ «ambiente globale, costituito in ultima istanza da informazioni» nel quale siamo immersi: L. FLORIDI, *The Logic of Information. A Theory of Philosophy as Conceptual Design*, Oxford, Oxford University Press, 2019, tr. it. di Massimo Durante, *Pensare l'infosfera. La filosofia come design concettuale*, Milano, Raffaello Cortina Editore, 2020, 13.

lungimiranti e flessibili, è, per altri versi, “nato già vecchio”. Del resto, dalla lettura dei documenti delle istituzioni dell’Unione europea redatti nel periodo di gestazione del GDPR emerge come tale impianto normativo sia stato messo a punto in un’epoca in cui a tali organismi non era ancora chiaro l’impatto che la protezione dei dati personali stava subendo per l’avvento dei *Big Data* e, di lì a poco avrebbe subito, per la diffusione dell’intelligenza artificiale (IA)<sup>110</sup>. Il lancio della strategia europea del *Digital Single Market*, la responsabilizzazione delle grandi piattaforme rispetto all’utilizzazione da parte di minori, l’incentivazione al riuso dei dati a scopi di ricerca scientifica, la necessità di assicurare l’affidabilità dei sistemi di IA, sarebbero stati fatti propri dalle Istituzioni dell’Unione Europea solo successivamente (in particolare, sotto la Commissione Juncker 2014-2019 e, poi, sotto la prima Commissione von der Leyen, 2019-2024, per continuare, tutt’ora, con la seconda)<sup>111</sup>.

Tali affermazioni saranno approfondite nel prosieguo, previe alcune necessarie precisazioni terminologiche attorno ai concetti di *Big Data* e IA, sì da perimetrire l’oggetto (seppur di “secondo grado”<sup>112</sup>) dell’elaborato.

### **Big Data e IA: pennellate definitorie**

Tra gli aspetti che accomunano i due fenomeni in parola vi sono sicuramente l’incertezza e relatività delle rispettive definizioni.

Consideriamo innanzitutto i *Big Data*: benché non esista una definizione condivisa, con questo termine si suole riferirsi al processo di analisi di “grandi” set di dati o, nell’accezione più diffusa, ai (“grandi”) set di dati stessi<sup>113</sup>. Il concetto è, in tutta evidenza, storicamente

---

<sup>110</sup> Si possono prendere ad esempio due documenti del Gruppo di lavoro art. 29 – organo consultivo che riuniva le autorità nazionali di vigilanza e protezione dei dati, poi sostituito dal Comitato europeo per la protezione dei dati con l’entrata in vigore del GDPR – nei quali è significativo che le parole *Big Data* e IA non compaiano affatto: il contributo congiunto assieme al Gruppo di lavoro su polizia e giustizia intitolato *The Future of Privacy*, 1° dicembre 2009 e il parere recante «*further input on the data protection reform discussions*», 5 ottobre 2012, n. 8. Il silenzio sarebbe stato rotto dal Garante europeo dei dati personali, con il parere del 16 marzo 2014, *Privacy and competitiveness in the age of big data*. In ogni caso, hanno tardato a venire a galla alcuni risvolti particolarmente problematici, come il rischio di distorsioni e discriminazioni, appena menzionato nel parere del Garante europeo 11 settembre 2015, n. 4, *Towards a new digital ethics Data, dignity and technology*, e del tutto assente, ad esempio, in un successivo parere del 23 settembre 2016, n. 8, in tema di «*coherent enforcement of fundamental rights in the age of big data*». Su questi profili, v. F. PIZZETTI, *GDPR, Codice novellato e Garante nell’epoca dei Big Data e della Intelligenza Artificiale*, in ID (a cura di), *Protezione dei dati personali in Italia tra GDPR e codice novellato*, Torino, Giappichelli, 2021, 233 ss.

<sup>111</sup> Tra gli interventi legislativi che hanno segnato il cambio di passo si ricordano il Data Governance Act, il Data Act, il Digital Market Act, il Digital Services Act e il Regolamento europeo sull’intelligenza artificiale. Per uno sguardo d’insieme e a ciascuna di queste discipline, cfr. F. PIZZETTI (a cura di), *La regolazione europea della società digitale*, Torino, Giappichelli, 2024.

<sup>112</sup> L’oggetto di studio sono, infatti, le norme che disciplinano le nuove tecnologie, mentre l’oggetto di secondo grado sono le tecnologie stesse.

<sup>113</sup> V., *ex multis*, S. FARO, N. LETTIERI, *Big Data: una lettura informatico-giuridica*, in AA.Vv. (a cura di), *Scritti per Luigi Lombardi Vallauri*, Padova, Cedam, 2016, I, 503. Va subito precisato che i *Big Data* non consistono necessariamente in dati personali, benché sia altamente probabile che almeno una parte di questi enormi set di dati riguardino persone fisiche identificate o identificabili ai sensi dell’art. 4, n. 1, GDPR. Se si accede a quella impostazione che distingue i concetti di dato e di informazione, i *Big Data* consistono in (grandi quantità di) dati, cioè rappresentazioni digitali di

condizionato: i *data* sono *big* quando, alla luce delle attuali conoscenze tecnologiche, sono difficili da analizzare, gestire, elaborare in un tempo ragionevole, a causa delle loro dimensioni e complessità. In assenza di soglie dimensionali predefinite, ciò che è considerato *big* oggi (proprio come ciò che è considerato intelligente, come vedremo) potrebbe non esserlo domani.

Non a caso i tentativi definitori hanno condotto all'elaborazione, oramai "classica", delle "V"<sup>114</sup>, che cercano di descrivere il fenomeno da un punto di vista qualitativo: alle tre "V" originarie – *volume*, *velocità* (ossia la rapidità del trattamento, ma anche della creazione dei dati, spesso generati in tempo reale e costantemente aggiornati) e *varietà* (delle fonti da cui i dati o i metadati provengono) – sono state affiancate una serie di altre caratteristiche, quali la *veridicità* o veracità (che, in realtà, ben potrebbe mancare laddove i *Big Data* siano costituiti da dati inesatti o da dati sintetici); il *valore* che è possibile estrarne (che dipende dalla loro natura, accuratezza, grandezza, etc.); la *vulnerabilità*, connessa al rischio di furto dei dati o di utilizzi illeciti (caratteristica che evidenzia la stretta relazione fra protezione dei dati e cybersicurezza); la *valenza* (ossia il grado di connessione del dato con altri dati); e, infine, la *visualizzazione* (cioè, la necessità di riassumere graficamente o comunque in maniera facilmente interpretabile le informazioni più rilevanti tratte da tali set di dati).

Tali tentativi definitori non sembrano cogliere nel segno: da un lato, la seconda e la terza "V" originarie sembrano costituire presupposti della prima (sono la generazione costante e in tempo reale e la provenienza dalle fonti più disparate a determinare l'enorme volume dei *Big Data*); dall'altro lato, le successive "V" parrebbero caratteristiche meramente accessorie o, comunque, conseguenziali all'aspetto dimensionale. Nel tornare, dunque, a quest'ultimo quale irrinunciabile punto di partenza, si può ipotizzare, come soglia dimensionale minima perché si possa parlare di *Big Data*, quella che rende necessario un trattamento di tipo automatizzato, grazie al quale realizzare analisi descrittive e, soprattutto, predittive.

Numerosi problemi definitori si riscontrano anche in relazione all'IA, non solo perché si tratta di un altro termine storicamente condizionato<sup>115</sup>, ma anche in quanto il dibattito attorno a tale concetto investe una pluralità di settori disciplinari – dalla psicologia cognitiva all'ingegneria informatica, dalla filosofia all'economia – e ha radici molto risalenti<sup>116</sup>.

---

fatti e atti, dai quali è possibile trarre sia informazioni che riguardano persone fisiche, identificate o identificabili (cioè, dati personali nel senso del GDPR) sia informazioni non riguardanti persone fisiche (cioè, dati non personali). S. CALZOLAIO, *op. cit.*, 22.

<sup>114</sup> Cfr., ad esempio, T. MAURO, *I Big Data tra protezione dei dati personali e diritto della concorrenza*, in R. PANETTA (a cura di) *Circolazione e protezione dei dati personali, tra libertà e regole del mercato*, Milano, Giuffrè, 2019, 652 s.

<sup>115</sup> Basti pensare che gli elettrocardiografi interpretativi furono accolti tra gli anni Settanta e Novanta come una delle conquiste dell'intelligenza artificiale in ambito medico; oggi queste tecnologie non sono certo ricomprese fra le tecniche di IA.

<sup>116</sup> Semplificando, si possono identificare tre posizioni, appresso sinteticamente riportate. Alcuni sottolineano la lunga tradizione filosofica, culturale e anche tecnologica dell'IA (dalle opere precorritrici di Alan Turing negli anni '40 al workshop al Dartmouth College del 1956). Altri identificano alcune particolari tecniche come iniziatrici dell'era della vera e propria IA, segnatamente quella particolare tipologia dell'apprendimento autonomo (*machine learning*), conosciuta come apprendimento profondo (*deep learning*): lo spartiacque sarebbe dunque nei primi anni 2000 o intorno al 2012, assurto ad anno-simbolo nella storia del *deep learning* per la vittoria della competizione ImageNet da

Il diritto positivo, quando si è trovato a dover definire e disciplinare l'IA, ne ha riconosciuto la natura di termine-ombrello, sotto il quale sono variamente ricondotti approcci e tecniche che creano sistemi automatizzati in grado, secondo la formula adottata dal legislatore europeo, di «funzionare con livelli di autonomia variabili», eventualmente adattandosi all'ambiente circostante, in grado di dedurre dai dati di ingresso che riceve (*input*), dati di uscita (*output*) «quali previsioni, contenuti, raccomandazioni o decisioni» (art. 3, n. 1, Regolamento sull'IA<sup>117</sup>).

Si tratta, nella maggior parte dei casi, di algoritmi<sup>118</sup> che sfruttano tecniche e approcci non basati su un modello (*non model-based*<sup>119</sup>): i sistemi di IA non sono, cioè, in grado di astrarre la regola causale sottostante il rapporto tra l'*input* e l'*output*, ma solamente di esprimere la probabilità che a un certo *input* corrisponda un certo *output*. Si pensi ai *Large Language Models* usati per l'IA generativa: l'algoritmo, in questi casi, non astrae ed applica una qualche regola o conoscenza, ma produce (genera, appunto) una lettera dopo l'altra sulla base della probabilità che, in un dato contesto, ad una segua l'altra (emulando, così, il linguaggio umano).

*Big Data* e IA sono inoltre fenomeni strettamente interconnessi: non è un caso che sia stato l'avvento dei primi, unitamente all'accresciuto potere computazionale, a permettere a tecniche già conosciute, ma relativamente poco sfruttate – come quelle dell'apprendimento autonomo – di dispiegare appieno le proprie potenzialità, generando crescente attenzione da parte della ricerca, del mercato, delle istituzioni.

I sistemi di IA hanno sollevato, tuttavia, una serie di problematiche specifiche e ulteriori, che possono schematicamente riassumersi ne: a) l'opacità “intrinseca”<sup>120</sup> dell'algoritmo di intelligenza artificiale, che spesso costituisce una vera e propria “scatola nera” (*black-box*), la cui “logica di funzionamento” consiste in centinaia di migliaia di parametri e correlazioni, dai quali non è tecnicamente possibile – specie nei sistemi non basati su un modello – astrarre

---

parte della rete neurale convoluzionale AlexNet, sviluppata da un gruppo di ricerca coordinato da Geoffrey Hinton (considerato l'ideatore della *backpropagation*). Altri ancora, infine – sullo sfondo della riflessione filosofica sul concetto di “intelligenza” e degli studi psicologici attorno alla tensione antropomorfizzante –, sostengono che neppure le tecniche più avanzate e che danno la parvenza di essere senzienti, come l'IA generativa, siano veramente “intelligenti” (e neppure propriamente “artificiali”, come ricorda il titolo della traduzione italiana – “Né intelligente, né artificiale” –, edita nel 2021 da Il Mulino, del volume di KATE CRAWFORD, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, Yale, Yale University Press, 2021).

<sup>117</sup> Regolamento (UE) 2024/1689 (Regolamento sull'intelligenza artificiale, più noto come *Ai Act*).

<sup>118</sup> Sulla differenza tra il concetto di algoritmo, che si riferisce a qualsiasi sequenza di istruzioni che genera un determinato risultato, e il concetto di intelligenza artificiale, si è espressa anche la giurisprudenza del Consiglio di Stato (v., ad es., Cons. St., sent. 25 novembre 2021, n. 7891).

<sup>119</sup> Cfr. P. TRAVERSO, *Breve introduzione tecnica all'Intelligenza Artificiale*, in *DPCE online*, n. 1/2022, 155 ss.

<sup>120</sup> Per distinguerla dagli altri due tipi di “scatola nera” osservabili con riferimento agli algoritmi di IA: quella “intenzionale” (dovuta alla secretazione del codice sorgente da parte di enti pubblici e privati secondo la disciplina del segreto commerciale o di stato) e quella “soggettiva” (dovuta alle conoscenze tecniche del soggetto medio, al quale risultano “opachi” sistemi che per i programmati sono del tutto o in buona misura intellegibili). La tripartizione è di J. BURREL, *How the machine ‘thinks’: Understanding the opacity in machine learning algorithms*, in *Big Data & Society*, 2016, 3 ss. Il contributo più famoso sul problema della “scatola nera” algoritmica si deve, però, a F. PASQUALE, *The Black Box society: The Secret Algorithms That Control Money and Information*, Cambridge, Harvard University Press, 2015.

alcuna regola che spieghi la produzione di un determinato *output*<sup>121</sup>; b) la graduale corrosione dell'autonomia umana da parte di sistemi che, da semplici strumenti, si trasformano, direttamente o indirettamente (per il graduale inconsapevole "affidamento" ai risultati della macchina)<sup>122</sup>, in veri e propri agenti; c) il problema delle distorsioni (*bias*) dei sistemi, che possono produrre "allucinazioni" (false informazioni) e generare discriminazioni (o, meglio, perpetrarle, atteso che i sistemi altro non fanno che riprodurre e, dunque, cristallizzare correlazioni e schemi esistenti nei dati di ingresso).

### GDPR: una normativa "nata vecchia"?

Tracciate queste brevi premesse introduttive, occorre riprendere la riflessione attorno alla duplice e opposta lettura del GDPR quale, da un lato, disciplina flessibile e, almeno nelle intenzioni, tesa a evitarne la rapida obsolescenza, e, dall'altro, come intervento incapace di cogliere le novità introdotte dai *Big Data* e più ancora dall'IA.

A un primo sguardo, l'impianto normativo sembra fondato su una serie di assunti che ne denunciano una certa staticità<sup>123</sup>: che sia possibile individuare un titolare del trattamento (o, al limite, più contitolari o titolari autonomi) e un soggetto responsabile ex art. 28 (eventualmente supportato da sub-responsabili), nonché un numero definito di interessati; che sia posto in essere un numero definito di trattamenti, ciascuno dei quali fondato su una precisa base giuridica; che sia possibile individuare *ex ante* le finalità per le quali i dati sono raccolti e trattati, considerato che ogni ulteriore trattamento è lecito solo se effettuato per finalità non incompatibili con quelle per le quali i dati sono stati raccolti (secondo il principio di limitazione delle finalità); che sia possibile raggiungere tutti gli interessati e renderli consapevoli delle caratteristiche dei trattamenti in corso attraverso un testo scritto (l'informativa).

Un quadro così delineato sembra male applicarsi alle problematiche sollevate dai *Big Data*: le sempre più ingenti quantità di dati raccolte, conservate, condivise fra moltissimi soggetti e il loro impiego per le finalità più diverse, creano "catene" di rapporti che difficilmente possono essere ricondotte al modello che fa capo a un titolare che tratta i dati di più interessati per finalità predeterminate. Per non parlare delle sfide precise dell'IA più sopra sinteticamente richiamate (opacità, autonomia, *bias*).

Per comprendere meglio i termini del problema, occorre tuttavia procedere a un'analisi più minuta, soffermandosi su alcuni profili specifici della disciplina disegnata dal GDPR, che se,

<sup>121</sup> Si segnala, a tal proposito, il filone di ricerca sulle intelligenze artificiali "spiegabili" (*Explainable AI*). Sembra tuttavia esistere un ineliminabile *tradeoff* tra spiegabilità e accuratezza dei sistemi, in base al quale quanto più un modello è semplice tanto più è facile da spiegare, ma presenta una resa più bassa, mentre per ottenere sistemi ad alte prestazioni occorre accettarne la complessità e, quindi, l'opacità. Sul punto, ad es., XU E AL., *Explainable AI: a brief survey on history, research areas, approaches and challenges*, in J. TANG E AL. (a cura di), *Natural Language Processing and Chinese Computing*, Cham, Springer, 2019, 565.

<sup>122</sup> Il fenomeno è anche conosciuto come "distorsione dell'automazione" (*automation bias*).

<sup>123</sup> Sul tema, cfr. F. PIZZETTI, *GDPR e codice novellato*, cit., 242.

per certi versi, hanno costituito un primo passo nella tutela dei diritti fondamentali nell’ambiente digitale, possono, tuttavia, essere interpretati come altrettanti punti di caduta della normativa in tema di protezione dei dati rispetto alle sfide poste dai fenomeni tecnologici in oggetto. Nel prosieguo saranno pertanto oggetto di approfondimento: 1) il concetto di dato personale, 2) la disciplina in tema di decisioni automatizzate e 3) la regolamentazione dei trattamenti di dati sensibili per l’addestramento degli algoritmi a fini di antidiiscriminatori.

### Crisi e resilienza del concetto di dato personale

La definizione del GDPR abbracciata dal legislatore europeo è ampia ed elastica: qualsiasi informazione riferibile a una persona «identificabile», anche «indirettamente» può costituire dato personale, in particolare attraverso il collegamento con un qualsiasi «identificativo», che può consistere, oltre che nel «nome» o in «un numero di identificazione» anche nei «dati relativi all’ubicazione» o in «uno o più elementi caratteristici della sua identità fisica, fisiologica, genetica, psichica, economica, culturale o sociale»<sup>124</sup>.

Un dato personale può perdere la caratteristica della “personalità” attraverso una serie di tecniche che abbiano come effetto l’anonimizzazione, cioè quel trattamento che – diversamente dalla mera pseudonimizzazione – comporta l’eliminazione irreversibile della possibilità di collegare i dati a una persona fisica.

Ebbene, nell’era dei *Big Data*, l’anonimizzazione in senso tecnico è, in molti casi, impossibile: attraverso il trattamento automatizzato e incrociato di grandi banche dati, spesso liberamente accessibili *online*, si è potuto, infatti, in un caso, identificare i soggetti ai quali si riferivano semplici punteggi attribuiti dagli utenti di una piattaforma di *streaming* ai contenuti fruiti, e, in un altro caso, ricongiungere a rispettivi interessati alcuni dati genetici resi pubblici a fini di ricerca<sup>125</sup>. Inoltre, così come è più labile il confine tra dato personale e non personale, anche la «distanza computazionale» che separa un dato personale comune da un

---

<sup>124</sup> Per comprendere questo punto è utile la comparazione con la disciplina statunitense: la nozione di *Personally Identifiable Information*, attorno alla quale ruota la (frastagliata) disciplina statunitense a tutela della riservatezza, prende in considerazione soltanto il dato riconducibile ad un soggetto identificato e non anche potenzialmente identificabile. Per questo si suole affermare che ogni *Personally Identifiable Information* sia dato personale, ma non viceversa. Cfr. R. DUCATO, *La crisi della definizione di dato personale nell’era del web 3.0. Una lettura civilistica in chiave comparata*, in F. CORTESE, M TOMASI (a cura di), *Le definizioni nel diritto*, Atti delle giornate di studio 30-31 ottobre 2015, Università degli Studi di Trento, 2016, 160.

<sup>125</sup> Si fa riferimento, nel primo caso, allo scandalo Netflix del 2006, nel secondo caso, invece, all’esperimento di re-identificazione condotto nel 2013 da un gruppo di ricercatori del MIT sui dati genetici dei partecipanti al progetto “1000 Genome”. Cfr., *ivi*, 162. Il tema ha avuto eco anche nell’Indagine conoscitiva sui *Big Data*, avviata nel 2017 e conclusasi nel 2020 dall’Autorità garante della concorrenza e del mercato, dall’Autorità per le garanzie nelle comunicazioni e dal Garante per la protezione dei dati personali.

dato appartenente alle categorie particolari di dati (ai sensi dell'art. 9 GDPR) è andata accorciandosi<sup>126</sup>.

Di fronte all'impossibilità di azzerare il rischio di re-identificazione, la nozione giuridica di anonimizzazione introduce elementi di flessibilità rispetto a quella tecnica: si considerano, cioè, anonimizzati (e, dunque, non personali) i dati quando sia diventato *ragionevolmente* impossibile ricollegarli a persone fisiche<sup>127</sup>. Il «test di ragionevolezza» si fonda su una serie di parametri oggettivi, tra i quali, il costo della re-identificazione (in termini economici e di tempo) e il valore dei dati in oggetto (a sua volta dipendente, ad esempio, dalla loro rarità, natura, volume), ma anche soggettivi (come le caratteristiche del potenziale terzo intruso, quali la disponibilità di risorse, l'accesso ad altre banche dati, etc.).

Nel valutare la probabilità che i mezzi di re-identificazione della persona siano effettivamente utilizzati, il considerando 26 GDPR<sup>128</sup> chiede di tenere in considerazione sia le «tecnologie disponibili al momento del trattamento» sia, con una valutazione prospettica, gli «sviluppi tecnologici»: anche l'anonimizzazione diventa, dunque, un concetto storicamente condizionato, rendendo relativo e dinamico il concetto di dato personale. Resta la domanda se il tentativo di evitare (o ritardare) l'obsolescenza della definizione di dato personale, agganciandola all'evoluzione tecnologica, sia di per sé sufficiente a garantire l'equilibrio tra la riservatezza e i molteplici altri diritti e interessi di rilievo costituzionale che con essa possano confliggere.

### **La disciplina delle decisioni basate unicamente su trattamenti automatizzati: opera incompiuta o ponte gettato verso l'era algoritmica?**

L'articolo 22 GDPR è sicuramente una delle disposizioni del Regolamento più innovative per il periodo in cui è stata redatta. Esso, come noto, disciplina i processi decisionali automatizzati, ponendo, al primo comma, un generale divieto di sottoporre gli interessati a un

<sup>126</sup> Tale distanza dipende da fattori oggettivi (tecnologie disponibili) e soggettivi (risorse economiche umane di cui un particolare soggetto potrebbe disporre per procedere alla re-identificazione), cfr. G. MALGIERI, G. COMANDÉ, *Sensitive-by-distance: quasi-health data in the algorithmic era*, in *Information & Communications Technology Law*, n. 3/2017, 229 ss. Uno studio di ricercatori di Stanford, ad esempio, ha dimostrato che è possibile inferire con sorprendente precisione lo stato di salute di una persona analizzando i metadati relativi alle conversazioni telefoniche (numeri digitati, durata della telefonata, etc.). Cfr. J. MAYER, P. MUTCHLER, J. C. MITCHELL, *Evaluating the privacy properties of telephone metadata*, in *PNAS*, n. 20/2016, 5536 ss.

<sup>127</sup> Più precisamente, il processo di anonimizzazione è tale quando impedisce a chiunque impieghi quell'insieme di dati anonimizzati, in combinazione con tutti i mezzi di cui può *ragionevolmente* disporre, di: 1) isolare una persona da un gruppo, 2) collegare un dato anonimizzato a dati personali contenuti in un altro insieme o 3) dedurre nuove informazioni su una persona a partire dal dato anonimizzato. Cfr. G. D'ACQUISTO, M. NALDI, *Big Data e privacy by design: anonimizzazione, pseudonimizzazione, sicurezza*, Torino, Giappichelli, 2017, 34.

<sup>128</sup> Il contenuto del considerando riprende un parere del Gruppo di lavoro articolo 29 sulle tecniche di anonimizzazione (parere n. 5/2014). Si segnala che il Comitato europeo per la protezione ha di recente espressamente richiamato i criteri per effettuare il «test di ragionevolezza» nel parere sull'uso dei dati di localizzazione e degli strumenti per il tracciamento dei contatti nel contesto della pandemia da Covid-19 (parere n. 4/2020).

trattamento *unicamente* automatizzato, compresa la profilazione<sup>129</sup>, quando questo produca «effetti giuridici» (ad esempio, il rifiuto di un permesso di soggiorno o della cittadinanza, la negazione di prestazioni previdenziali, l'annullamento di un contratto) o «incida in modo analogo significativamente» sulla persona (ad esempio, influenzando l'ammissibilità al credito, negandole opportunità di impiego o l'accesso all'università)<sup>130</sup>; al secondo comma, contiene ben tre eccezioni al divieto, di un'ampiezza tale da svuotarlo largamente della sua portata garantista<sup>131</sup>.

Nei casi in cui sia ammessa la decisione unicamente automatizzata, l'interessato ha diritto di ottenere l'intervento umano da parte del titolare del trattamento, di esprimere la propria opinione e di contestare la decisione. Egli ha altresì diritto, non solo di sapere dell'esistenza del processo automatizzato, ma di ottenere anche, nell'ambito della consueta informativa e in caso di esercizio del diritto di accesso ai dati, informazioni significative sulla logica utilizzata, nonché sull'importanza e sulle conseguenze di tale trattamento<sup>132</sup>.

Ora, poiché i sistemi di intelligenza artificiale eseguono trattamenti automatizzati di dati (anche) personali, che in molti casi si traducono in profilazione e/o decisioni automatizzate, si può certamente affermare che l'art. 22 costituisca il primo nucleo di garanzie della persona rispetto ai sistemi di IA. Tale tutela è, tuttavia, parziale, per almeno tre ordini di ragioni.

In primo luogo, l'ambito di applicazione è, per un verso, più ampio, e, per un altro verso, più ristretto, atteso che il concetto di «processo decisionale automatizzato» non coincide con quello di «sistema di IA»: ben possono configurarsi, da un lato, decisioni automatizzate che non si basano su algoritmi propriamente rientranti nelle tecniche di intelligenza artificiale (è il caso dell'algoritmo per l'assegnazione delle sedi di lavoro agli insegnanti della scuola pubblica, oggetto delle pronunce del Consiglio di Stato delle quali si tratterà nel prosieguo) e, dall'altro lato, sistemi di IA che non comportano alcuna decisione che produca effetti giuridici o incida in modo analogo significativamente sull'interessato (si pensi ai un robot ad uso industriale, o a generatori di testo e di immagini<sup>133</sup>).

---

<sup>129</sup> Le Linee guida del Gruppo di lavoro art. 29 sul processo decisionale automatizzato relativo alle persone fisiche e sulla profilazione, 3 ottobre 2017, spec. 8 s., chiariscono i rapporti tra le due nozioni: da un lato, la profilazione (cioè, a mente dell'art. 4, n. 4 GDPR, «qualsiasi forma di trattamento automatizzato di dati personali consistente nell'utilizzo di tali dati personali per valutare determinati aspetti personali relativi a una persona fisica, in particolare per analizzare o prevedere aspetti riguardanti il rendimento professionale, la situazione economica, la salute, le preferenze personali, gli interessi, l'affidabilità, il comportamento, l'ubicazione o gli spostamenti di detta persona fisica») può non condurre ad alcuna decisione automatizzata (se rimane, ad esempio, uno strumento di valutazione del posizionamento strategico dell'azienda sul mercato); viceversa, una decisione automatizzata non per forza segue a un'attività di profilazione (si pensi a una multa irrogata a seguito delle rilevazioni di un autovelox).

<sup>130</sup> Gli esempi si ispirano al considerando 71 GDPR e alle sopra citate Linee guida, 24.

<sup>131</sup> Il trattamento unicamente automatizzato è comunque lecito: *a)* se è stato ottenuto il consenso (libero, informato ed esplicito) dell'interessato; *b)* se è autorizzato dal diritto dell'Ue o degli Stati Membri; *c)* se è necessario per la conclusione di un contratto.

<sup>132</sup> Cfr. artt. 13, co. 2, lett. *f*), 14, co. 2, lett. *g*) e 15, co. 1, lett. *h*).

<sup>133</sup> Più complesso il discorso, ad esempio, in caso di robot chirurgici, per i quali potrebbero comunque prospettarsi interpretazioni ampliative della nozione di decisione che produca «effetti che incidono in modo analogo significativamente sull'interessato». Nella giurisprudenza della Corte di giustizia dell'Unione europea, del resto, è rinvenibile qualche apertura in direzione di un'interpretazione estensiva di tale concetto: si pensi alla nota sentenza

In secondo luogo, le sopra menzionate garanzie di trasparenza poste dal GDPR (fornire al soggetto interessato informazioni significative sulla logica utilizzata, così da permettergli, eventualmente, di esercitare il diritto di contestare la decisione ed esprimere la propria opinione) presentano alcuni limiti testuali, che si è cercato di superare in via interpretativa, dando luogo a un vivace dibattito dottrinale sull'esistenza, l'ambito di applicazione e il contenuto di un "diritto alla spiegazione" della decisione automatizzata<sup>134</sup>.

Il dibattito verte, innanzitutto, sulla riferibilità di tali garanzie a *tutti* i trattamenti automatizzati o solamente a quelli *unicamente* automatizzati (che escludano, cioè, ogni coinvolgimento umano). Secondo una lettura restrittiva del combinato disposto degli artt. 13, co. 2, lett. f), 14, co. 2, lett. g) e 15, co. 1, lett. h) e del comma 1 e dell'art. 22 GDPR, la previsione un coinvolgimento umano, anche minimo, escluderebbe l'applicabilità delle tutele. Non manca, tuttavia, chi ha suggerito di estenderle ad ogni decisione automatizzata, ancorché sottoposta a revisione umana, attraverso una lettura ampliativa fondata sul considerando 71 o direttamente sugli artt. 13 e 14. Da parte sua, il Gruppo di lavoro articolo 29, nelle Linee guida sulla profilazione, ha, da un lato, confermato il diverso regime di trasparenza per le decisioni fondate su trattamenti unicamente automatizzati rispetto agli altri casi, nei quali fornire informazioni significative sulla logica utilizzata non costituisce un obbligo, bensì una «buona prassi»<sup>135</sup>; dall'altro lato, riconoscendo il rischio di coinvolgimenti umani finti per aggirare gli obblighi di trasparenza rafforzati, ha richiesto che il contributo umano sia «significativo e non costituisca un semplice gesto simbolico» e che sia effettuato «da una persona che dispone dell'autorità e della competenza per modificare la decisione». Il problema si sposta dunque sulla determinazione del *quantum* di coinvolgimento umano che escluda l'applicazione delle maggiori tutele.

Altra questione controversa riguarda il contenuto dell'obbligo informativo: se, cioè, il GDPR a) fondi un diritto alla spiegazione della singola decisione imponendo al titolare, ad esempio, di fornire un quadro chiaro di come i dati in ingresso si relazionino alle predizioni; oppure, b) stabilisca un (più limitato) diritto a ricevere informazioni *ex ante* sul funzionamento del sistema e sull'importanza e le conseguenze del trattamento, ma non un diritto ad una spiegazione *ex post* della singola decisione<sup>136</sup>; o, ancora, c) ponga, al di là di quella che si riduce, agli occhi di alcuni, ad una diatriba sul *nomen* (diritto alla spiegazione o ad essere informati), un diritto il cui contenuto è circoscritto dalla sua natura *strumentale* all'esercizio del diritto di contestare la decisione (pertanto tutte e solo le informazioni necessarie a tale

---

Schufa (Causa C-634/21, SCHUFA Holding (Scoring), 7 dicembre 2023) che vi ha incluso i sistemi di *credit scoring*, utilizzati per stimare la solvibilità di una persona attraverso il calcolo del tasso di probabilità che questa sia in grado di onorare impegni di pagamento in futuro.

<sup>134</sup> Il contributo che ha dato avvio al dibattito è: B. GOODMAN, S. FLAXMAN, *European Union Regulations on Algorithmic Decision Making and a "Right to Explanation"*, in *AI Magazine*, n. 3/2017, 50 ss.

<sup>135</sup> Pur restando fermo, anche in questo caso, l'obbligo per il titolare di fornire informazioni, espresse in forma chiara e semplice, sull'esistenza del trattamento automatizzato e sulle sue conseguenze. Cfr. Linee guida sulla profilazione (v. nt. 26), 28.

<sup>136</sup> S. WACHTER, B. MITTELSTADT, L. FLORIDI, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, in *International Data Privacy Law*, n. 2/2017, 92.

scopo saranno oggetto dell’obbligo informativo)<sup>137</sup>; infine, *d)* ponga al centro il soggetto e la sua capacità di comprensione, richiedendo al titolare di rendere intelligibile<sup>138</sup> (cioè, a un tempo, conoscibile e comprensibile), il processo decisionale automatizzato.

In terzo e ultimo luogo, le medesime garanzie trasparenza si scontrano con un dato fattuale, di natura tecnica, ineluttabile: come ricostruire la logica di una decisione se l’algoritmo che l’ha assunta o suggerita consiste in una “scatola nera”? Se, infatti, neppure lo sviluppatore del sistema è in grado di comprendere la *ratio* sottostante ai risultati generati, tanto meno, si osserva, potrà farlo il soggetto (lo Stato, l’istituto di credito, etc.) che impiega l’algoritmo per assumere la decisione.

Se, per le regioni sopra esposte, l’art. 22 non pare offrire garanzie sufficienti a rispondere a tutte le sfide poste dai sistemi intelligenti, si deve tuttavia riconoscere che le disposizioni sulle decisioni automatizzate e i relativi obblighi informativi hanno costituito terreno fertile per interpretazioni ampliative da parte della giurisprudenza, delle Autorità di controllo e della dottrina.

In particolare, nell’ambito di strumenti a supporto o in sostituzione di provvedimenti amministrativi, la tutela accordata dalla giurisprudenza ha finito per eccedere i confini “naturali” della tutela della riservatezza e della protezione dei dati in senso stretto, per ricollegarsi ad altre garanzie di rango sovranazionale (come il diritto a una “buona amministrazione”, di cui all’art. 42 CDFUE) e costituzionale (il principio di buon andamento e imparzialità di cui all’art. 97 Cost. e corollari, come il principio di trasparenza e l’obbligo di motivazione dei provvedimenti, che si ricollegano anche al diritto di difesa in giudizio di cui all’art. 24 Cost.). Così argomentando, il Consiglio di Stato<sup>139</sup> è giunto a enucleare i tre principi che emergono «in termini generali dal diritto sovranazionale»<sup>140</sup> con riferimento al rapporto tra potere amministrativo e mezzi informatici, principi che tentano di dare risposta ai tre problemi, rispettivamente, dell’opacità, dell’autonomia umana e dei *bias*, richiamati alla fine del § 2: i) il principio di conoscibilità (dell’esistenza di processi decisionali automatizzati da parte del destinatario), che si completa con il principio di comprensibilità (che fonda il diritto a ricevere informazioni significative sulla logica utilizzata, corredate di necessità di spiegazioni che traducano la «formula tecnica» nella «regola giuridica» ad essa sottesa); ii) il principio di non esclusività della decisione algoritmica, in base al quale deve esistere all’interno del processo decisionale un contributo umano in grado di controllare, validare ovvero smentire la decisione, secondo il paradigma c.d. *human in the loop* (la portata di tale

---

<sup>137</sup> L’accento sull’aspetto teleologico è posto da A. D. SELBST, J. POWLES, *Meaningful information and the right to explanation*, in *International Data Privacy Law*, n. 4/2017, 233 ss.

<sup>138</sup> Trad. dell’a. per il termine inglese *legibility*, sul quale v. G. MALGIERI, G. COMANDÉ, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, in *International Data Privacy Law*, n. 4/217, 253 e *passim*.

<sup>139</sup> Si v. la “storica” sent. Cons. St. n. 2270 del 2019 e le cd. tre sentenze gemelle Cons. St. Sez. VI nn. 8472, 8473, 8474 del 2019.

<sup>140</sup> Sono parole della sent. Cons. St., n. 8472/2019.

principio non è, tuttavia, pacifica<sup>141</sup>); iii) il principio di non discriminazione algoritmica, in base al quale devono essere messe in atto misure adeguate a prevenire il rischio di errori della macchina: in particolare devono essere modificati i dati in ingresso, sì da evitare che i risultati proposti dall'algoritmo siano *distorti*<sup>142</sup>.

Tali principi sono stati richiamati, in alcune occasioni, anche dal Garante per la protezione dei dati<sup>143</sup>, la cui “giurisprudenza” era stata, del resto, anticipatrice delle tutele introdotte con l’art. 22<sup>144</sup>. La letteratura giuspubblicistica, infine, ha spesso ripreso tali principi all’interno della sempre più ampia riflessione sui sistemi intelligenti. Benché il punto di partenza di tale riflessione sia sempre la necessità di lasciare orientare l’IA dal diritto costituzionale<sup>145</sup>, i punti di arrivo dei singoli autori e “scuole” sono diversificati. Alcuni, ad esempio, hanno enucleato “nuovi diritti” di rilievo sostanzialmente costituzionale (come il diritto alla spiegazione, il diritto ad essere resi consapevoli della natura umana o artificiale dell’interlocutore, il diritto alla non esclusività della decisione algoritmica)<sup>146</sup> o “nuovi principi” (come il principio di sorveglianza umana)<sup>147</sup>. Altri hanno suggerito di applicare analogicamente principi tratti da ambiti differenti (come il principio di precauzione, sorto in materia ambientale<sup>148</sup>). Altri ancora hanno adottati approcci “settoriali”, che individuano le sfide peculiari poste dalle

<sup>141</sup> A tale riguardo, si è registrata una diversità di vedute tra il Consiglio di Stato e i giudici di primo grado in ordine alla estensione del necessario apporto umano rispetto a quello algoritmico: secondo i giudici di primo grado alle procedure informatiche può essere attribuito solamente «un ruolo strumentale e meramente ausiliario in seno al procedimento amministrativo e giammai dominante o surrogatorio dell’attività dell’uomo»; il Consiglio di Stato, invece, si è limitato a vietare procedimenti amministrativi unicamente automatizzati, senza esigere una particolare incidenza dell’apporto umano rispetto alla decisione finale. Sulle divergenze e consonanze di questi filoni giurisprudenziali, v., per tutti, R. FERRARA, *Il giudice amministrativo e gli algoritmi. Note estemporanee a margine di un recente dibattito giurisprudenziale*, in *Dir. amm.*, n. 4/2019, 773 ss.

<sup>142</sup> Il punto è esplicitato nel Regolamento solamente nel considerando 71, ma, come vedremo, può leggersi come declinazione del principio generale dell’Ue di non discriminazione.

<sup>143</sup> Che li ha ripresi esplicitamente in più occasioni (v. ad esempio, il Parere al Consiglio di Stato sulle nuove modalità di ripartizione del fondo sanitario tra le regioni proposte dal Ministero della salute e basate sulla stratificazione della popolazione, del 5 marzo 2020, e il punto 4 del *Decalogo per la realizzazione di servizi sanitari nazionali attraverso sistemi di Intelligenza Artificiale*, pubblicato nel settembre 2023).

<sup>144</sup> Sul punto si può leggere l’efficace sintesi di G. CERRINA FERONI, *Intelligenza artificiale e ruolo della protezione dei dati personali. L’analisi di Ginevra Cerrina Feroni*, reperibile sul sito del Garante (doc. web. 9855742).

<sup>145</sup> Non di rado inteso, in senso esplicito e non, come non ristretto all’ambito nazionale: un “costituzionalismo” inteso come «quel modello di tutela delle situazioni giuridiche soggettive, così come si è diffuso nelle culture dei paesi occidentali, ove ha trovato una sintesi nelle Carte dei diritti dell’uomo, e dal quale si va diffondendo in tutte le parti del globo terraqueo». B. CARAVITA DI TORITTO, *Principi costituzionali e intelligenza artificiale*, in U. RUFFOLO (a cura di), *Intelligenza artificiale, Il diritto, i diritti, l’etica*, Milano, Giuffrè, 2020, 461.

<sup>146</sup> V., ex multis, C. CASONATO, *Costituzione e intelligenza artificiale: un’agenda per il prossimo futuro*, in *BioLaw Journal*, Fascicolo Speciale n. 2/2019, 711 ss.; L. RINALDI, *Intelligenza artificiale, diritti e doveri nella Costituzione italiana*, in *BioLaw Journal*, n. 1/2022, 215 ss.

<sup>147</sup> M. FASAN, *I principi costituzionali nella disciplina dell’Intelligenza Artificiale. Nuove prospettive interpretative*, in *BioLaw Journal*, 1/2022, 181 ss.

<sup>148</sup> A. SIMONCINI, *L’algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *BioLaw Journal*, n. 1/2019, 86 ss. In senso solo parzialmente analogo, si è insistito sulla necessità che il diritto assuma «un’impronta intergenerazionale», che «senta il dovere di pensare al futuro», di «mantenere nel tempo le condizioni essenziali della convivenza umana»: A. D’ALOIA, *Il diritto verso “il mondo nuovo”. Le sfide dell’Intelligenza Artificiale*, in *BioLaw Journal*, n. 1/2019, 30.

caratteristiche dei sistemi di IA nei singoli contesti: si pensi ai riflessi dei sistemi di supporto alla decisione giudiziaria sulla discrezionalità del giudice e sul principio di non vincolatività del precedente<sup>149</sup>; o ai sistemi di supporto alla diagnosi e alla scelta terapeutica che vengano a inserirsi nella relazione di cura, ponendo, tra le altre cose, il problema di garantire il principio del consenso informato ai trattamenti sanitari<sup>150</sup>; o ai sistemi di sorveglianza biometrica, laddove è evidente che la riservatezza, in quella particolare declinazione che è l'anonymato in luogo pubblico, mostra un carattere di strumentalità all'esercizio del diritto di manifestazione del pensiero e del diritto di riunione<sup>151</sup>.

In sintesi, l'art. 22 GDPR, pur con tutti i limiti che derivano dall'estensione delle eccezioni al divieto di trattamenti unicamente automatizzati, dalle ristrettezze dell'ambito oggettivo di tutela e dalla difficoltà sia di interpretazione estensiva e sia di applicazione pratica dei requisiti di trasparenza, rappresenta un vero e proprio ponte fra la disciplina dei dati personali e le risposte normative che legislatori, giurisprudenza e dottrina stanno cercando di dare alle problematiche poste dai sistemi di IA.

#### **La tensione tra riservatezza e non discriminazione algoritmica, tra GDPR e Ai Act**

Discorrendo dei limiti dell'art. 22, si è già potuta osservare l'insufficienza della tutela apprestata con riguardo all'esigenza che siano adottate, durante la fase di addestramento degli algoritmi di IA, misure tali da mitigare le distorsioni dei sistemi (alle quali ci si riferisce comunemente con il termine inglese “*bias*”) che possano provocare errori nei risultati e discriminazioni<sup>152</sup>.

L'accuratezza e non discriminatorietà di un sistema di IA è soggetta al principio “*garbage in-garbage out*” (c.d. GIGO): esse dipendono, cioè, in larga parte, dalla quantità e qualità dei dati di addestramento, che devono essere, tra le altre cose, sufficientemente rappresentativi, per evitare il fenomeno, molto frequente, per cui il sistema funziona bene con i dati utilizzati in

---

<sup>149</sup> V., per tutti, M. LUCIANI, *La decisione giudiziaria robotica*, in *Rivista AIC*, n. 3/2018, 872.

<sup>150</sup> L. SCAFFARDI, *La medicina alla prova dell'Intelligenza Artificiale*, in *DPCE online*, n. 1/2022, 349 ss.; D. MORANA, T. BALDUZZI, F. MORGANTI, *La salute “intelligente”: eHealth, consenso informato e principio di non-discriminazione*, in *federalismi.it*, n. 34/2022, 127 ss.

<sup>151</sup> Cfr. G. MOBILIO, *Tecnologie di riconoscimento facciale. Rischi per i diritti fondamentali e sfide regolative*, Napoli, Editoriale Scientifica, 2021, e, se si vuole, il commento all'art. 5 dell'Ai Act (in particolare le voci: *I sistemi biometrici vietati (1): riconoscimento delle emozioni e categorizzazione biometrica; I sistemi biometrici vietati (2): identificazione biometrica remota “in tempo reale”; Allegato II; Allegato III*), redatto da chi scrive per il *Commentario all'Ai Act* a cura di A. Mantelero, G. Resta e G. Riccio ed edito da WoltersKluwer, in corso di pubblicazione.

<sup>152</sup> Per comprendere la possibilità che si cumulino errori e discriminazioni, è utile l'oramai classico esempio del sistema per il riconoscimento del carcinoma alla pelle: da un lato, la presenza di *bias* nei set di dati di addestramento (perché, ad esempio, insufficientemente rappresentativi di soggetti dalla pelle scura) può determinare errori nella diagnosi; dall'altro, la maggiore frequenza di “falsi positivi” e “falsi negativi” con riferimento a certe etnie potrebbe costituire una forma di discriminazione: se il sistema fosse impiegato all'interno del Servizio sanitario nazionale, ben si potrebbe parlare di un *vulnus* al principio di equità dell'accesso alle cure.

fase di sviluppo, ma presenta un basso rendimento in contesti reali<sup>153</sup>. Nel riprodurre correlazioni esistenti tra i dati di ingresso, inoltre, i sistemi possono perpetrare o esacerbare discriminazioni esistenti, specie quando i risultati forniti dall'algoritmo sono utilizzati per prendere decisioni che riguardano persone fisiche<sup>154</sup>. È dunque necessario condurre analisi approfondite sugli algoritmi per verificarne la non discriminatorietà ed eventualmente procedere a correzioni. Tali operazioni comportano una serie di rischi: in molti casi, infatti, non basta escludere le informazioni sull'appartenenza a categorie protette per evitare che gli algoritmi producano risultati discriminatori, occorrendo bensì procedere a correggere “attivamente” l'algoritmo, trattando, a tal fine, dati sensibili<sup>155</sup>.

Di fronte a tale scenario, alcuni hanno cercato di fondare la liceità o addirittura l'obbligatorietà dei trattamenti finalizzati a individuare e correggere le distorsioni sul testo del GDPR. In particolare, si è prospettata una lettura del principio di esattezza dei dati e del principio di correttezza del trattamento alla luce del considerando 71<sup>156</sup>, il quale esplicitamente prevede «al fine di garantire un trattamento corretto e trasparente nel rispetto dell'interessato» che sia «opportuno» che il titolare del trattamento: a) metta in atto «misure tecniche e organizzative adeguate al fine di garantire, in particolare, che siano rettificati i fattori che comportano inesattezze dei dati e sia minimizzato il rischio di errori»; b) impedisca «effetti discriminatori nei confronti di persone fisiche sulla base della razza o dell'origine etnica, delle opinioni politiche, della religione o delle convinzioni personali, dell'appartenenza sindacale, dello status genetico, dello stato di salute o dell'orientamento sessuale».

Con analoghi risultati, si è sostenuta la necessità che la valutazione del rischio secondo l'approccio della protezione dei dati fin dalla progettazione (*privacy by design*) prenda in considerazione anche i potenziali effetti discriminatori. In altre parole, i trattamenti antidiscriminazione farebbero parte di quelle «misure tecniche e organizzative adeguate» volte a «tutelare i diritti degli interessati»<sup>157</sup> che il titolare è obbligato a porre in essere ex art. 25 GDPR<sup>158</sup>. Sarebbe, dunque, proprio l'art. 25 il «ponte gettato verso il futuro», poiché

---

<sup>153</sup> Si tratta del fenomeno del c.d. “*overfitting*”.

<sup>154</sup> A prescindere che il procedimento possa essere ricondotto nell'alveo delle decisioni automatizzate ex art. 22, per le quali, come detto, occorre, tra l'altro, verificare la decisione produca effetti giuridici o che incidono in modo analogo significativamente sulla persona.

<sup>155</sup> Lo sottolineava già l'Agenzia dell'Unione europea per i diritti fondamentali (FRA) nel Focus #BigData: *Discrimination in data-supported decision making*, pubblicato il 29 maggio 2018, 8.

<sup>156</sup> Sul valore non precettivo dei considerando cfr. la *Guida pratica comune del Parlamento europeo, del Consiglio e della Commissione per la redazione dei testi legislativi dell'Unione europea*, 2015, 31 s.

<sup>157</sup> È quindi evidente che la nozione di “interessato”, in questa prospettiva, deve essere interpretata estensivamente, sì da ricoprendere non solo le persone già interessate o coloro che il titolare sa che diventeranno interessate, bensì tutti coloro i cui dati potranno essere sottoposti al trattamento da parte dell'algoritmo che il titolare (nell'esempio, lo sviluppatore del *software*) sta elaborando. Cfr. F. PIZZETTI, *La protezione dei dati personali e la sfida dell'Intelligenza Artificiale*, in Id, *Intelligenza Artificiale, protezione dei dati personali e regolazione*, Torino, Giappichelli, 2018, 127.

<sup>158</sup> Sul punto convergono la posizione delle istituzioni dell'Unione europea (espressa, ad esempio, dal FRA, v. nt. 51) e quella del Consiglio d'Europa (v. le *Linee guida sulla protezione degli individui con riferimento al trattamento dei dati personali nel mondo dei Big Data*, del 23 gennaio 2017).

imporrebbbe di «inserire, nei programmi e negli algoritmi che costituiscono il linguaggio della macchina, dei principi etici che aiutino le macchine stesse a decidere»<sup>159</sup>.

Trattare l'art. 25 come una sorta di clausola generale ha il sicuro vantaggio di rendere la disciplina più flessibile e meno esposta al rischio di obsolescenza. Ne risente, tuttavia, la certezza e l'uniforme applicazione del diritto, essendo sostanzialmente demandata all'interprete – cioè, ai titolari stessi<sup>160</sup>, prima, ed, eventualmente, alle autorità di controllo e giurisdizionali, poi – l'individuazione degli specifici obblighi ai quali il titolare è tenuto per assicurare l'adesione al principio di *privacy by design*. In assenza di un'interpretazione evolutiva, condivisa e costante nel senso sopra richiamato delle lacunose disposizioni del GDPR, la normativa sulla protezione dei dati personali sembra insufficiente a rispondere alle esigenze di contemperamento tra le istanze della riservatezza e quelle della non discriminazione. Il Garante sul punto si è mostrato prudente e non ha fornito, per il momento, indicazioni specifiche sui trattamenti a fini antidiscriminatori, pur richiamando, come già accennato, gli approdi della giurisprudenza del Consiglio di Stato sul principio di non discriminazione algoritmica<sup>161</sup>.

Il problema della prevenzione del *bias* può leggersi come contemperamento tra il principio di minimizzazione, secondo il quale i dati personali trattati devono essere adeguati, pertinenti e limitati a quanto necessario rispetto alle finalità, e il principio di non discriminazione, principio generale dell'Unione al quale la Corte di Giustizia ha accordato efficacia diretta anche nei rapporti orizzontali<sup>162</sup>, che impone di svolgere ogni operazione necessaria al fine di scongiurare il rischio di discriminazione. Conferme di questa tensione e un tentativo di risoluzione, con l'imposizione di obblighi più specifici agli sviluppatori e l'esplicitazione di una deroga al divieto di trattamento di dati sensibili ex art. 9 GDPR è stata raccolta dal legislatore dell'Unione con il Regolamento sull'IA.

In particolare, l'art. 10 impone agli sviluppatori di valutare se i propri sistemi di IA ad alto rischio presentano distorsioni che possano avere un impatto negativo sui diritti fondamentali o comportare discriminazioni e di adottare le misure volte a prevenirle e attenuarle (par. 2, lett. f) e g)). Qualora, poi, sia strettamente necessario per rilevare e rimuovere le distorsioni e ricorrano tutte le condizioni di cui alle lett. a-f) del par. 5, i fornitori possono eccezionalmente trattare anche categorie particolari di dati personali ex art. 9 GDPR, fatta salva l'adozione di

---

<sup>159</sup> *Ivi*, 129.

<sup>160</sup> Titolari che, in virtù del principio di responsabilizzazione, sono chiamati non semplicemente ad attuare una serie di requisiti formali, bensì a individuare e a dimostrare di aver predisposto le garanzie più indicate nel caso concreto e da un punto di vista sostanziale a proteggere i diritti degli interessati. Cfr. L. CALIFANO, *Il Regolamento UE 2016/679 e la costruzione di un modello uniforme di diritto europeo alla riservatezza e alla protezione dati personali*, in EAD, C. COLAPIETRO (a cura di), *Innovazione tecnologica e valore della persona. Il diritto alla protezione dei dati personali nel Regolamento UE 2016/679*, Napoli, Editoriale Scientifica, 2017, 34.

<sup>161</sup> L'Autorità belga per la protezione dei dati, invece, ha sostenuto che le operazioni atte a individuare e correggere le distorsioni discendono dal principio di correttezza del trattamento e accuratezza dei dati in una brochure informativa pubblicata a dicembre 2024, *Artificial Intelligence Systems and the GDPR A Data Protection Perspective*.

<sup>162</sup> V., ad esempio, Corte di Giustizia, sent. 19 aprile 2016, C-441/14, *Dansk Industri (DI)*.

tutele adeguate per i diritti e le libertà fondamentali. Tra le condizioni che devono essere soddisfatte vi è la circostanza che il rilevamento e rimozione non possano essere efficacemente realizzati tramite dati sintetici o anonimizzati. Semplificando, la normativa sembra suggerire che nel bilanciamento tra minimizzazione e non discriminazione prevale la seconda (dunque i *bias* devono sempre essere individuati ed eliminati), ma nella scelta delle tecniche per effettuare tali operazioni si deve optare per quelle meno invadenti possibile dal lato della riservatezza.

Il legislatore europeo non ha dunque confidato negli spazi interpretativi lasciati dal GDPR per fondare la liceità e, ancor più, l’obbligatorietà dell’adozione di misure antidiscriminazione che implicano il trattamento di dati personali comuni e sensibili, preferendo adottare ulteriori e specifiche disposizioni. Si è trattato di un passaggio importante nella ricerca del punto di equilibrio tra minimizzazione e non discriminazione, nel tentativo di coordinare e integrare GDPR e Regolamento sull’IA. Tentativo, peraltro, secondo alcuni, fallito o, al limite, rimasto incompiuto, auspicandosi ulteriori chiarificazioni da parte della Commissione e del Comitato per la protezione dei dati<sup>163</sup>.

### **Una (brevisima) conclusione**

L’analisi condotta conferma l’ipotesi iniziale di questo contributo: alcuni dei nodi normativi del GDPR maggiormente esposti all’evoluzione tecnologica – che si tratti delle fondamenta stesse della disciplina in oggetto (come il concetto di dato personale), o di tentativi di tutela integrale dell’“umano” rispetto ai rischi dell’automazione (la garanzie che circondano le decisioni automatizzate), o, ancora, di aspetti che impongono una visione sistemica degli ordinamenti eurounitario e nazionale (la base giuridica per i trattamenti di dati in funzione anti-discriminazione algoritmica) – oscillano tra obsolescenza e lungimiranza.

E, con essi, oscilla anche il ruolo del giurista, chiamato, in un primo momento, a individuare le soluzioni ermeneutiche che permettano alla normativa di rispondere in modo efficace alle sfide dell’evoluzione tecnologica e, solo in un secondo momento, a valutare interventi legislativi mirati per colmare le lacune residue o attenuarne le rigidità.

### **Bibliografia**

BECK U., *The Metamorphosis of the World*, Cambridge, Polity Press, 2016, tr. it. di Marco Cupellaro, *La metamorfosi del mondo*, Bari, Editori Laterza, 2017.

BURREL J., *How the machine ‘thinks’: Understanding the opacity in machine learning algorithms*, in *Big Data & Society*, 2016.

---

<sup>163</sup> Così, ad esempio, per il servizio studi del Parlamento europeo, S. DE LUCA, M. FEDERICO, *Algorithmic discrimination under the AI Act and the GDPR*, European Parliamentary Research Office, febbraio 2025.

CALIFANO L., *Il Regolamento UE 2016/679 e la costruzione di un modello uniforme di diritto europeo alla riservatezza e alla protezione dati personali*, in EAD, C. COLAPIETRO (a cura di), *Innovazione tecnologica e valore della persona. Il diritto alla protezione dei dati personali nel Regolamento UE 2016/679*, Napoli, Editoriale Scientifica, 2017.

CALZOLAIO S., *Vulnerabilità della società digitale e ordinamento costituzionale dei dati*, in *Rivista italiana di informatica e diritto*, n. 2/2023.

CARAVITA DI TORITTO B., *Principi costituzionali e intelligenza artificiale*, in U. RUFFOLO (a cura di), *Intelligenza artificiale. Il diritto, i diritti, l'etica*, Milano, Giuffré, 2020.

CASONATO C., *Costituzione e intelligenza artificiale: un'agenda per il prossimo futuro*, in *BioLaw Journal*, Fascicolo Speciale n. 2/2019.

CERRINA FERONI G., *Intelligenza artificiale e ruolo della protezione dei dati personali. L'analisi di Ginevra Cerrina Feroni*, reperibile sul sito del Garante (doc. web. 9855742).

D'ACQUISTO G., NALDI M., *Big Data e privacy by design: anonimizzazione, pseudonimizzazione, sicurezza*, Torino, Giappichelli, 2017.

D'ALOIA A., *Il diritto verso "il mondo nuovo". Le sfide dell'Intelligenza Artificiale*, in *BioLaw Journal*, n. 1/2019.

DE LUCA S., FEDERICO M., *Algorithmic discrimination under the AI Act and the GDPR*, European Parliamentary Research Office, febbraio 2025.

DUCATO R., *La crisi della definizione di dato personale nell'era del web 3.0. Una lettura civilistica in chiave comparata*, in F. CORTESE, M TOMASI (a cura di), *Le definizioni nel diritto*, Atti delle giornate di studio 30-31 ottobre 2015, Università degli Studi di Trento, 2016.

FARO S., LETTIERI N., *Big Data: una lettura informatico-giuridica*, in AA.Vv. (a cura di), *Scritti per Luigi Lombardi Vallauri*, Padova, Cedam, 2016.

FASAN M., *I principi costituzionali nella disciplina dell'Intelligenza Artificiale. Nuove prospettive interpretative*, in *BioLaw Journal*, 1/2022.

FERRARA R., *Il giudice amministrativo e gli algoritmi. Note estemporanee a margine di un recente dibattito giurisprudenziale*, in *Dir. amm.*, n. 4/2019.

FLORIDI L., *The Logic of Information. A Theory of Philosophy as Conceptual Design*, Oxford, Oxford University Press, 2019, tr. it. di Massimo Durante, *Pensare l'infosfera. La filosofia come design concettuale*, Milano, Raffaello Cortina Editore, 2020.

GOODMAN B., FLAXMAN S., *European Union Regulations on Algorithmic Decision Making and a "Right to Explanation"*, in *AI Magazine*, n. 3/2017.

LUCIANI M., *La decisione giudiziaria robotica*, in *Rivista AIC*, n. 3/2018.

MALGIERI G., COMANDÉ G., *Sensitive-by-distance: quasi-health data in the algorithmic era*, in *Information & Communications Technology Law*, n. 3/2017.

MALGIERI G., COMANDÉ G., *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, in *International Data Privacy Law*, n. 4/217, 253 e *passim*.

MAURO T., *I Big Data tra protezione dei dati personali e diritto della concorrenza*, in R. PANETTA (a cura di) *Circolazione e protezione dei dati personali, tra libertà e regole del mercato*, Milano, Giuffrè, 2019.

MAYER J., MUTHLER P., MITCHELL J. C., *Evaluating the privacy properties of telephone metadata*, in *PNAS*, n. 20/2016.

MOBILIO G., *Tecnologie di riconoscimento facciale. Rischi per i diritti fondamentali e sfide regolative*, Napoli, Editoriale Scientifica.

MORANA D., BALDUZZI T., MORGANTI F., *La salute “intelligente”: eHealth, consenso informato e principio di non-discriminazione*, in *federalismi.it*, n. 34/2022.

PASQUALE F., *The Black Box society: The Secret Algorithms That Control Money and Information*, Cambridge, Harvard University Press, 2015.

PIZZETTI F. (a cura di), *La regolazione europea della società digitale*, Torino, Giappichelli, 2024.

PIZZETTI F., *GDPR, Codice novellato e Garante nell’epoca dei Big Data e della Intelligenza Artificiale*, in ID (a cura di), *Protezione dei dati personali in Italia tra GDPR e codice novellato*, Torino, Giappichelli, 2021.

PIZZETTI F., *La protezione dei dati personali e la sfida dell’Intelligenza Artificiale*, in ID, *Intelligenza Artificiale, protezione dei dati personali e regolazione*, Torino, Giappichelli, 2018.

RINALDI L., *Intelligenza artificiale, diritti e doveri nella Costituzione italiana*, in *BioLaw Journal*, n. 1/2022.

SCAFFARDI L., *La medicina alla prova dell’Intelligenza Artificiale*, in *DPCE online*, n. 1/2022.

SELBST A. D., POWLES J., *Meaningful information and the right to explanation*, in *International Data Privacy Law*, n. 4/2017.

SIMONCINI A., *L’algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *BioLaw Journal*, n. 1/2019.

SIMONCINI A., SUWEIS S., *Il cambio di paradigma nell’intelligenza artificiale e il suo impatto sul diritto costituzionale*, in *Rivista di filosofia del diritto*, n. 1/2019.

TRAVERSO P., *Breve introduzione tecnica all’Intelligenza Artificiale*, in *DPCE online*, n. 1/2022.

WACHTER S., MITTELSTADT B., FLORIDI L., *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, in *International Data Privacy Law*, n. 2/2017.

XU E AL., *Explainable AI: a brief survey on history, research areas, approaches and challenges*, in TANG J. E AL. (a cura di), *Natural Language Processing and Chinese Computing*, Cham, Springer, 2019.

## **Le sfide regolatorie della videosorveglianza diffusa tra innovazione tecnologica e tutela dei diritti fondamentali - di *Marco Ladu*, Professore associato di diritto costituzionale e pubblico presso l'Università degli Studi e-Campus**

La videosorveglianza<sup>164</sup>, generalmente intesa quale tecnica di osservazione e registrazione sistematica di spazi, persone o comportamenti, si basa oggi sull'applicazione di una serie di tecnologie avanzate che – nella c.d. Digital Era – divengono emblematiche del rapporto tra individuo ed esercizio di poteri di controllo da parte di soggetti pubblici e privati. Essa, inoltre, può rappresentare, congiuntamente, uno strumento di adeguata prevenzione e uno strumento arbitrario di oppressione: il c.d. occhio elettronico – che si caratterizza per una protrazione a oltranza dell'atto di osservare (cui si aggiunge la capacità di archiviazione di quanto captato) – sottrae quest'ultima attività dai limiti fisiologici della condotta umana ed espone al rischio che autorità pubbliche o soggetti privati impieghino i sistemi di videosorveglianza per finalità illecite o comunque ultronelle rispetto a quelle prefissate e ritenute legittime per via normativa.

Anzitutto per tale ragione, se da un lato la videosorveglianza può contribuire in misura significativa alla tutela della sicurezza pubblica e privata<sup>165</sup>, dall'altro si configura quale fatto che incide direttamente su diritti e libertà fondamentali, riconosciuti e protetti a livello costituzionale – ben oltre il diritto alla riservatezza<sup>166</sup> – il che implica un'analisi giuridica rigorosa del fenomeno sotto un altro tipo di occhio, quello ‘vigile’ del costituzionalista.

Tanto premesso, se si tiene conto della crescente proliferazione degli impianti di videosorveglianza, della capacità di dispiegare attività di controllo su ampia scala e dell'integrazione con le più disparate tecnologie (id est, ad esempio, quella di riconoscimento biometrico) non si può prescindere dal considerare la videosorveglianza come un fatto rilevante in misura sempre crescente sotto il profilo giuridico-costituzionale poiché, oggi, all'evoluzione delle tecnologie dovrebbe (ma, spesso e volentieri, non è così) corrispondere la capacità del sistema giuridico di rispondere al ‘nuovo’ con presidi giuridici adeguati.

---

<sup>164</sup> Oltre ai riferimenti bibliografici presenti nel resto del contributo, si rinvia per tutti, per una prima ricostruzione efficace del tema della videosorveglianza, ai lavori raccolti nel volume M. MANETTI, R. BORRELLO (a cura di), *Videosorveglianza e privacy*, Firenze, Angelo Pontecorbo Editore, 2010.

<sup>165</sup> Per una prima ricostruzione in chiave costituzionale del concetto di ‘sicurezza’ si rinvia, invece, alle riflessioni di G. DE VERGOTTINI, *La difficile convivenza fra libertà e sicurezza: la risposta delle democrazie al terrorismo. Gli ordinamenti nazionali*, in *Atti del XVIII Convegno Annuale dell'Associazione* (Bari, 17-18 ottobre 2003), Padova CEDAM, 2007, 56 ss.; M. DOGLIANI, *Il volto costituzionale della sicurezza*, in Astrid Rassegna, n. 22/2010 e A. PACE, *Libertà e sicurezza. Cinquant'anni dopo*, in *Diritto e Società*, n. 2/2013, 177 ss.

<sup>166</sup> Si nota, infatti, un'elevata attenzione generalizzata rispetto alla tutela della privacy e della riservatezza, ma allo stesso tempo una minore sensibilità rispetto alla tutela dei diritti fondamentali, che pure sono direttamente incisi da pratiche come la videosorveglianza. Con il presente contributo si intende riportare l'attenzione sulle problematiche di ordine strettamente costituzionale, che si intersecano certamente con la tutela della privacy e della riservatezza, ma che rappresentano un segmento molto più ampio di analisi che non deve essere trascurato. I diritti coinvolti in tema di videosorveglianza sono molti e diversificati e devono assumere un peso specifico al fine di assicurare tutela effettiva alla dignità e alla libertà della persona.

L'impressione di chi scrive è, difatti, quella che l'attuale normativa in materia di videosorveglianza<sup>167</sup> rischi di non riuscire più a prevenire e contenere lesioni di diritti e libertà fondamentali che si stanno facendo lentamente strada, man mano che sul mercato vengono resi disponibili, a prezzi accessibili ed estremamente concorrenziali, strumenti avanzati di controllo e captazione audio-video; precisamente, la normativa sulla privacy applicabile a tali attività, così come i correlati controlli, rischiano di non essere sufficientemente adeguati a proteggere dalle situazioni di fatto che oggigiorno vengono delineandosi in misura crescente.

In quest'ottica, il presente contributo si propone di esplorare da un lato i limiti e le garanzie costituzionali della videosorveglianza – con l'intento di abbozzare un possibile ‘statuto giuridico’ della videosorveglianza costituzionalmente compatibile – e dall'altro di offrire qualche spunto di riflessione su questioni aperte e criticità che restano sullo sfondo e che, tuttavia, meritano di essere urgentemente affrontate per approntare nuovi presidi e nuove forme di tutela e/o, quantomeno, rafforzare gli strumenti di garanzia esistenti rendendoli efficaci ed effettivi.

Entrando nel merito, seppur in sintesi, dei presupposti costituzionali della videosorveglianza, se è vero che non esiste una norma costituzionale che menziona espressamente tale attività, essa può trovare legittimazione in diverse previsioni costituzionali, fra le quali spiccano quelle concernenti la tutela della sicurezza, della proprietà privata, dell'iniziativa economica e dei suoi corollari (ivi compresa la tutela del c.d. patrimonio aziendale) nonché dell'amministrazione della giustizia.

In primo luogo, infatti, occorre richiamare – come diretta espressione di una delle modalità di estrinsecazione della dignità e della libertà personale (artt. 2 e 13 Cost.) – l'articolo 16 della Costituzione italiana, il quale, riconoscendo la libertà di circolazione e di soggiorno a ciascun cittadino ne ammette una limitazione, sorretta da una riserva rinforzata di legge, per i soli motivi di sanità e di sicurezza. Anche se non espressamente richiamati dalla Costituzione e anche se non direttamente collegati a previsioni attuative della riserva di legge appena richiamata, è innegabile che la videosorveglianza si pone comunque in diretta correlazione con la libertà del cittadino di circolare e soggiornare liberamente su tutto il territorio nazionale poiché la pienezza di tale libertà subisce un ineluttabile contemperamento con l'esigenza di sicurezza. La ‘libera’ circolazione e il ‘libero’ soggiorno, da questo punto di vista, subiscono, difatti, non tanto una limitazione in senso proprio – perché il cittadino, in presenza di sistemi di videosorveglianza, non viene privato o limitato rispetto al diritto di soggiornare e circolare sul territorio –, ma una rimodulazione in quanto, in presenza di sistemi di videosorveglianza, il

<sup>167</sup> In linea generale, il quadro normativo italiano in materia di videosorveglianza non risulta né univoco né sistematico e ciò in quanto accanto alla disciplina generale sul trattamento dei dati personali (D.lgs. 196/2003 prima e Regolamento (UE) 2016/679 poi) e al provvedimento del Garante per la protezione dei dati personali dell'8 aprile 2010, convivono disposizioni di settore, in particolare connesse alla tutela della sicurezza urbana (si vedano, ad esempio, le normative relative alla videosorveglianza nei Comuni o sui trasporti pubblici). Sebbene non agevole, la ricostruzione del quadro normativo trova comunque dei punti fermi nei principi di proporzionalità, finalità, minimizzazione e trasparenza del trattamento, oltre che nelle specifiche garanzie previste per l'ambito lavorativo dallo Statuto dei lavoratori (art. 4 L. 300/1970). Cfr. *Videosorveglianza* (voce), in L. BOLOGNINI, E. PELINO (diretto da), *Codice della disciplina privacy*, Milano, Giuffrè Francis Lefebvre, 2024, 1367 ss.

cittadino è comunque indotto ad agire in un determinato modo, ad esempio prestando maggiore attenzione a quel che fa e a come lo fa. Si tratta, osservando il fenomeno dal punto di vista giuridico, dell’attuazione di meccanismi che dovrebbero indurre ad un rafforzamento della capacità della persona di autodisciplinarsi nel rispetto delle regole, ma che – senza tema di smentita – finiscono in ogni caso per far assumere una diversa forma alle modalità di esercizio delle libertà di cui all’articolo 16 della Costituzione (e di altre libertà collegate, di cui si dirà).

In secondo luogo, l’installazione di sistemi videosorveglianza – specie quelli nel diretto controllo dei soggetti privati – trovano una copertura costituzionale negli articoli 41 e 42 della Costituzione e ciò in quanto, rispettivamente, «La proprietà privata è riconosciuta e garantita dalla legge, che ne determina i modi di acquisto, di godimento e i limiti allo scopo di assicurarne la funzione sociale e di renderla accessibile a tutti» (art. 42, c. 2, Cost.) e la libera iniziativa economica privata se pur «libera» (art. 41, c. 1, Cost.), «Non può svolgersi in contrasto con l’utilità sociale o in modo da recare danno alla salute, all’ambiente, alla sicurezza, alla libertà, alla dignità umana» (art. 41, c. 2, Cost.). Leggendo unitariamente i testi citati, è possibile affermare, in senso generale, che la videosorveglianza non solo non si pone in contrasto con suddette disposizioni, ma – a ben vedere, concretamente osservando – può concorrere a rafforzare l’esigenza di garantire sicurezza ogni qual volta essa è predisposta a tutelare la proprietà privata o, ad esempio, il patrimonio aziendale<sup>168</sup>.

Tutto ciò trova un rafforzamento, a monte, nell’articolo 14 della Costituzione, al cui primo comma è sancita l’inviolabilità del domicilio, il che giustifica modalità di controllo e vigilanza che mirino al rispetto di tale diritto fondamentale.

Ulteriormente, la videosorveglianza può assumere un’importante funzione preventiva nell’ambito della giustizia, trovando così un ulteriore fondamento negli articoli 24, 101, 111 e 112 della Costituzione, che garantiscono il diritto alla tutela giurisdizionale (art. 24 Cost.), stabiliscono la soggezione del giudice soltanto alla legge (art. 101, c. 2, Cost.), lo svolgimento di un giusto processo (art. 111 Cost.) nonché l’obbligo di esercizio dell’azione penale da parte del Pubblico Ministero (art. 112 Cost.). In tale prospettiva, prendendo tali principi quale espressione ‘emblematica’ del sistema giustizia (ma altre norme costituzionali ben potrebbero concorrere ad ampliare il novero), occorre tenere conto del fatto che le immagini raccolte tramite sistemi legittimi di videosorveglianza possono essere determinati sia in fase di indagine – ben potendo costituire elemento probatorio sia a carico che a discarico del soggetto sottoposto alle indagini – sia in fase decisoria, divenendo elemento di prova oggetto di valutazione del giudice ai fini della decisione.

---

<sup>168</sup> Discorso a parte meriterebbe il controllo esercitato da parte del datore di lavoro che si giustifica con la necessità di tutelare l’organizzazione e il patrimonio aziendale, ma che deve essere pur sempre bilanciato con il rispetto della dignità del lavoratore e con i limiti imposti dalla legge e, in particolare, dallo Statuto dei lavoratori.

Ed è doveroso notare, infine, che la videosorveglianza si pone in termini di compatibilità/incompatibilità anche a livello sovranazionale<sup>169</sup>, onde assumono particolare rilievo gli articoli 5 e 8 della CEDU, i quali sanciscono, rispettivamente, il diritto alla libertà e alla sicurezza (art. 5 CEDU) e il diritto al rispetto della vita privata e familiare (art. 8 CEDU)<sup>170</sup>.

Anche se richiamate per brevitas, le suddette disposizioni costituzionali e sovranazionali si pongono, dunque, quali principi e/o diritti e/o doveri di rango costituzionale che, se non rendono l'attività di videosorveglianza incompatibile con la Costituzione latamente intesa, dall'altro lato ne indicano più o meno esplicitamente i limiti.

Ecco, pertanto, perché è indispensabile procedere con un delicato bilanciamento delle disposizioni costituzionali: l'attività di videosorveglianza, se pur legittima e coperta sotto il profilo costituzionale, deve sempre esplicarsi nel rispetto pedissequo di una molteplicità di altri diritti e libertà, tra cui certamente la dignità umana (art. 2 Cost.), la libertà personale (art. 13 Cost.), la libertà e inviolabilità del domicilio (art. 14 Cost.), la segretezza della corrispondenza e di ogni altra forma di comunicazione (art. 15), la libertà di circolazione e di soggiorno (art. 16 Cost.), la libertà di riunione (art. 17 Cost.), la libera manifestazione del pensiero (art. 21 Cost.), ma anche la tutela della dignità sui luoghi di lavoro (artt. 4 e 35 Cost.).

La raccolta d'immagini e suoni, a fini di sicurezza, non è dunque mai un atto che si può considerare ‘neutro’ o ‘asettico’ nei contesti di democrazia costituzionale, poiché implica tanto l'esercizio di un potere di osservazione a tutela di alcune prerogative costituzionalmente protette, quanto la necessità che del suo esercizio sia fatto un uso compatibile con altrettanti importanti diritti e libertà, al fine specifico che essa non si traduca mai in una forma, più o meno latente, di ‘controllo sociale’ da parte dell'autorità pubblica e/o da parte dei privati cittadini.

Ecco perché, oltre al bilanciamento tra disposizioni costituzionali, è indispensabile che l'uso della videosorveglianza rispetti – con una verifica caso per caso<sup>171</sup> – anche il principio di proporzionalità, al fine di assicurare al cittadino non solo la tutela della propria privacy, ma anche il pedissequo rispetto dei diritti e delle libertà fondamentali, che non devono subire eccessive compressioni o anomale limitazioni. In definitiva, sul punto, è tanto vero che la libertà non è assoluta e illimitata quanto è vero che essa non deve subire contrazioni ingiustificate o, analogamente, giustificate in nome di una concezione di sicurezza e tutela della sfera privata individuale che esorbita dall'alveo costituzionale.

---

<sup>169</sup> Significativa, in merito, l'adozione delle fondamentali *Linee guida 3/2019 sul trattamento dei dati personali attraverso dispositivi video* adottate dall'*European Data Protection Board* (EDPB) il 29 gennaio 2020 (versione 2.0).

<sup>170</sup> Per un approfondimento si veda, tra gli altri, S. BAROLE, P. DE SENA, V. ZAGREBELSKY, *Commentario breve alla Convenzione europea dei diritti dell'uomo e delle libertà fondamentali*, Padova, CEDAM, 2012.

<sup>171</sup> In questo senso, rispetto al bilanciamento tra diritti in tema di sicurezza, T. F. GIUPPONI, *La sicurezza e le sue “dimensioni” costituzionali*, in *Forum di Quaderni Costituzionali*, 2008, 19 (anche in anche in S. VIDA (a cura di), *Diritti umani: trasformazioni e reazioni*, Bologna, Bononia University Press, 2008, 275-301) secondo il quale «mentre l'individuazione dell'oggetto di tutela di una libertà è operazione generale e astratta, l'operatività delle specifiche delimitazioni eventualmente previste e, ancora di più, l'eventuale bilanciamento con altri interessi costituzionali appaiono operazioni da attuare concretamente e caso per caso».

## **La funzione della videosorveglianza pubblica: qualche rilievo sulla recente introduzione delle bodycam in dotazione alle Forze di polizia ex art. 21, d.l. 48/2025 (c.d. decreto sicurezza)**

È bene spendere qualche parola più generale rispetto all'attività di videosorveglianza a seconda che essa sia svolta da autorità pubbliche o da soggetti privati, poiché cambiano non soltanto gli adempimenti, ma anche le forme di responsabilità e gli scopi correlati alla natura del soggetto che dispone dell'apparato di videosorveglianza.

Nel caso in cui la videosorveglianza sia esercitata da autorità pubbliche, infatti, essa diviene immediata espressione dell'esercizio del potere statale, con la necessità che siano rispettati i principi di legalità, necessità e proporzionalità. Tali attività, dunque, essendo svolte dallo Stato per la tutela di preminenti interessi pubblici, devono poggiarsi su esplicite previsioni legali, essere svolte nel rispetto del bilanciamento tra i diritti costituzionalmente protetti (fra tutti, la tutela della sicurezza pubblica e il contrasto alla criminalità), limitarsi allo stretto necessario, essere proporzionate rispetto alle finalità che concretamente persegono, senza che si giunga ad un contesto di vero e proprio ‘panoptismo’, in cui ciò il cittadino, nella consapevolezza di poter essere costantemente osservato, non è mai in grado di distinguere tempi, modi e spazi in cui la videosorveglianza diviene funzionale alla tutela di specifici e dichiarati pubblici interessi.

Ecco perché gli impianti di videosorveglianza predisposti dalle autorità pubbliche non dovrebbero mai risolversi in tecnologie in grado di controllare a tappeto gli spazi pubblici<sup>172</sup> perché, inevitabilmente, questo tipo di controlli fuoriesce – a parere di chi scrive – da un ambito di tollerabilità costituzionale. Quando la sorveglianza viene esercitata da autorità pubbliche, il ‘test’ di compatibilità costituzionale deve diventare stringente: in gioco non vi è solo la libertà della persona, ma il rapporto genetico tra individuo e potere.

In questa prospettiva, risulta centrale l'applicazione del principio di stretta legalità, tale per cui le attività di videosorveglianza pubblica, impattando in misura significativa sulla libertà della persona, secondo me dovrebbero: a) essere stabilite da previsioni legislative di diretta promozione degli organi parlamentari; b) delimitate, sempre da previsioni legislative di promozione parlamentare, rispetto ai tempi, alle modalità e alle finalità; c) soggette a verifiche periodiche e sistematiche da parte di organismi e autorità indipendenti.

Per le ragioni sopra esposte, esprimo la mia preoccupazione per l'introduzione nel nostro ordinamento, peraltro proprio attraverso lo strumento della decretazione d'urgenza (sic!), di recenti misure volte a dotare le Forze di polizia di bodycam.

Il riferimento è al decreto-legge 11 aprile 2025, n. 48 (c.d. decreto sicurezza), contenente “Disposizioni urgenti in materia di sicurezza pubblica, di tutela del personale in servizio, nonché di vittime dell'usura e di ordinamento penitenziario”, al cui primo comma dell'articolo

---

<sup>172</sup> Si pensi, al riguardo, alla recente proposta della sindaca di Settimo Torinese, Elena Piastra, la quale ha proposto l'impiego di droni aerei per il controllo e la sicurezza del territorio. In particolare, si veda G. GROSSO, *Sicurezza: «Droni per sorvegliare la città». L'idea contro vandali e ladri*, articolo del 30 aprile 2024 disponibile all'indirizzo [www.torinocronaca.it](http://www.torinocronaca.it).

21 – rubricato “Dotazione di videocamere al personale delle Forze di polizia” – si prevede che «il personale delle Forze di polizia impiegato nei servizi di mantenimento dell’ordine pubblico, di controllo del territorio e di vigilanza di siti sensibili nonché in ambito ferroviario e a bordo dei treni può essere dotato di dispositivi di videosorveglianza indossabili, idonei a registrare l’attività operativa e il suo svolgimento».

Tale disposizione nel più ampio contesto del menzionato decreto, assai criticato in dottrina<sup>173</sup>, è passibile di alcuni rilievi.

Un primo rilievo ritengo sia da porre in stretta relazione la ‘leggerezza’ politica e ideologica con la quale vengono affrontate tematiche estremamente rilevanti nel bilanciamento tra rispetto delle libertà della persona e tutela della sicurezza pubblica e privata. Tali norme, infatti, vengono adottate fuori dal confronto parlamentare, con un decreto-legge in assenza dei presupposti costituzionali (ed è quasi superfluo spendere parole, dopo decenni di abuso, sull’illegittimità costituzionale di tale prassi). In particolare, quest’ultima circostanza è gravemente confermata, utilizzando le parole del Consiglio direttivo dell’Associazione Italiana dei Professori di Diritto Penale (AIPDP), dal fatto che oggi si registra – rispetto all’introduzione del decreto sicurezza – «l’anomalo ricorso alla decretazione d’urgenza in materia penale per trasferire in un decreto-legge un intero disegno di legge presentato oltre un anno fa e al cui esame sono state dedicate un centinaio di sedute tra Camera e Senato, con l’audizione di numerosi professori ed esperti. Il decreto-legge viene così impropriamente utilizzato come un disegno di legge ad effetto immediato, creando un precedente che potrebbe alimentare una prassi che svilisce il ruolo del Parlamento»<sup>174</sup>.

Un secondo rilievo riguarda il fatto che la misura in questione resta comunque generica e si limita a stabilire che sia possibile dotare le Forze di polizia di tali dispositivi, senza prevedere casi in cui forse sarebbe il caso, anche a tutela degli agenti di polizia, di rendere obbligatorio l’utilizzo delle bodycam e, in ogni caso, senza stabilire tempi, modalità e limiti stringenti al loro utilizzo.

Occorre, infine, riflettere sulla duplice funzione delle bodycam: se da un lato, infatti, esse tutelano le Forze di polizia nel loro operato (si pensi alla complessità di certe operazioni che quotidianamente vengono svolte sui mezzi di trasporto o in contesti cittadini denotati da un elevato tasso di criminalità) dall’altro garantiscono i cittadini poiché rendono trasparenti e verificabili le modalità con le quali le Forze di polizia agiscono, in virtù della delicatezza del ruolo ricoperto dalle stesse.

Il decreto sicurezza, pertanto, prevede sì disposizioni che sembrano concorrere ad un incremento della sicurezza nel nostro Paese, ma a ben vedere contiene, nel caso specifico

---

<sup>173</sup> Solo a titolo di esempio, E. GROSSO, *È compito della Repubblica. Note sul DDL Sicurezza*, in *Giustizia Insieme*, 4 novembre 2024; M. RUOTOLI, *Su alcune criticità costituzionali del c.d. pacchetto sicurezza (A.S. 1236)*, in *Sistema Penale*, 9 ottobre 2024. Ma anche iniziative a livello di studiosi e docenti di materie giuridiche, tra cui quelle ricordate da F. Q., *Appello contro il decreto Sicurezza, il testo integrale e i nomi dei firmatari: da Zagrebelsky a De Siervo*, in *Il Facto Quotidiano*, 27 aprile 2025; *Sul “pacchetto sicurezza” varato con decreto-legge*, 9 aprile 2025, 1 (documento dell’Associazione italiana dei professori di diritto penale - AIPDP) disponibile sul portale [www.questionejustizia.it](http://www.questionejustizia.it)

<sup>174</sup> *Sul “pacchetto sicurezza” varato con decreto-legge*, cit., 1.

delle bodycam, solo una disposizione generica che apre ad una possibilità, senza fare chiarezza e senza dare effettività alle esigenze di sicurezza.

Il decreto sicurezza, purtroppo, crea meno sicurezza e «sacrifica la sicurezza dei diritti»<sup>175</sup>, alimentando il focolaio della mera propaganda politica a scopi elettorali.

### **La proliferazione della videosorveglianza privata e la necessità di estendere i controlli sul rispetto della normativa vigente**

Per quel che attiene, invece, alla videosorveglianza nel contesto privatistico, essa si configura come esercizio della libertà del soggetto di tutelare i propri beni e la propria integrità. Tuttavia, essa non può sfuggire al vaglio costituzionale poiché, come si è detto, il nostro ordinamento giuridico impone che anche i rapporti intersoggettivi siano conformi ai precetti costituzionali e che quindi siano sempre rispettati i principi di dignità, libertà e riservatezza, in virtù del principio di efficacia orizzontale dei diritti fondamentali (la c.d. Drittewirkung indiretta tedesca).

Cambiando punto di vista, ritengo di dover precisare che, se è pur vero non spetta al privato cittadino farsi carico della tutela della sicurezza pubblica, non è possibile negare il fatto che il legittimo controllo esercitato dai singoli privati possa contribuire, nel complesso, a garantire una maggiore sicurezza collettiva (e, in particolare, la c.d. sicurezza urbana<sup>176</sup>).

Il cittadino che installa un impianto di videosorveglianza diventa, di fatto, un soggetto capace di incidere sulla sfera giuridica altrui e, di conseguenza, questa attività deve ritenersi legittima solo se conforme, in primis, alle disposizioni costituzionali e, soprattutto, solo se strettamente rispettosa del principio di proporzionalità. In concreto, chi installa impianti di videosorveglianza a tutela degli spazi esterni alla propria abitazione o alla propria attività imprenditoriale (e sono queste le ipotesi più frequenti) deve limitarsi a riprendere solo ed esclusivamente gli spazi immediatamente adiacenti (ad esempio, l'ingresso dell'abitazione, del capannone o del garage), ma non deve riprendere spazi comuni (per esempio pubbliche vie o piazze) se non per le porzioni strettamente necessarie a consentire la ripresa dello spazio da tutelare e non deve interferire con la vita privata di altri, eccedendo rispetto alla finalità dichiarata. Inoltre, il privato è tenuto ad informare, con apposita cartellonistica, della presenza

---

<sup>175</sup> Così autorevolmente E. DOLCINI, *Un Paese meno sicuro per effetto del decreto-legge sicurezza*, in *Sistema Penale*, 15 maggio 2025, il quale esprime un «giudizio critico che già [gli] era stato suggerito da approccio di insieme, concentrato sulle linee di fondo del provvedimento. Il d.l. sicurezza rappresenta un nuovo attacco ai principi fondamentali dello Stato di diritto» poiché «L'enfasi che il d.l. 48/2025 pone sul diritto alla sicurezza sacrifica la sicurezza dei diritti».

<sup>176</sup> Il tema della sicurezza urbana – la cui nozione è ben ricostruita da A. PAJNO, *Alla ricerca della nozione di "sicurezza urbana"*, in *Astrid online*, 10 settembre 2008 – è stata oggetto di un significativo dibattito in Italia, alimentato da alcune proposte che puntavano ad introdurre le cc.dd. ronde cittadine. In particolare, con disposizioni urgenti in materia di sicurezza delle città, il decreto-legge n. 14 del 2017 (poi convertito con modificazioni dalla legge 18 aprile 2017, n. 48) ha modificato la disciplina consentendo ai comuni di installare telecamere di videosorveglianza per fini di sicurezza urbana. La definizione di sicurezza urbana è però ibrida poiché interseca sia aspetti legati al decoro urbano sia legati alla prevenzione di minacce all'incolmabilità pubblica, entrando nel campo della sicurezza pubblica.

di impianti di videosorveglianza specificando: a) da chi è effettuata la registrazione e chi funge da responsabile del trattamento dei dati acquisiti; b) il termine entro il quale le registrazioni saranno conservate; c) la finalità della videosorveglianza; d) un recapito al quale rivolgersi per ottenere l'accesso ai propri dati e poter esercitare gli altri diritti connessi riconosciuti dalla normativa in vigore<sup>177</sup>.

Nonostante tali previsioni, bisogna constatare, specie in relazione all'installazione ‘fai da te’ di impianti di videosorveglianza, che spesso e volentieri i privati non ottemperano agli adempimenti stabiliti. Non è infrequente, infatti, accorgersi della presenza di telecamere che intercettano porzioni di abitazioni private e ampi segmenti di spazi pubblici solo ad una attenta osservazione, in assenza di qualsivoglia cartellonistica che dovrebbe essere presente e persino a posta in evidenza e a distanza tale da preannunciare che di lì a poco si transiterà in una zona soggetta a videosorveglianza.

Si profila, dunque, la necessità di forme di controllo capillari che diano effettività alle previsioni normative in vigore, altrimenti le tutele e le garanzie rischiano di restare lettera morta. In questo, ritengo che sia determinante – oltre a un serio intervento parlamentare – il ruolo dei giudici e, soprattutto, del Garante per la protezione dei dati personali che, come le altre autorità amministrative indipendenti, svolge funzioni di regolazione e di controllo poste a tutela di interessi pubblici e privati di natura costituzionale, assicurando elevate competenze tecnico-giuridiche.

### **Qualche osservazione conclusiva de iure condendo**

Per concludere, intendo guardare, in una prospettiva de iure condendo, alla dimensione che sta via via assumendo il fenomeno della videosorveglianza. Si tratta, infatti, di un sistema ormai ad altissime prestazioni, che si serve di complessi e sempre più avanzati strumenti di analisi automatizzata.

Tra le tecniche che possiamo considerare maggiormente intrusive spicca certamente quella del riconoscimento biometrico, che consente di affiancare alla mera captazione audio-video la capacità di individuare esattamente il soggetto che transita e/o sosta in una determinata area ed è oggetto della captazione. In questo caso specifico, quindi, ad emergere prepotentemente è l'assenza di anonimato in una fase in cui la videosorveglianza dovrebbe unicamente attestarsi sul piano della prevenzione. L'individuazione del soggetto, infatti, dovrebbe derivare da successivi accertamenti autorizzati preventivamente dall'autorità giudiziaria per scopi legati all'amministrazione della giustizia e, in particolare, all'accertamento di comportamenti meritevoli di approfondimento in quanto, con un certo grado di probabilità, potrebbero essere riconducibili a fattispecie illecite.

---

<sup>177</sup> Si rinvia, per le disposizioni e le indicazioni di dettaglio, a quanto indicato sotto la voce ‘videosorveglianza’ sul portale dell’Autorità Garante per la Protezione dei Dati Personalni.

Ad aumentare esponenzialmente la capacità di riuscita della combinazione delle diverse tecnologie è, poi, l'intelligenza artificiale, la cui applicazione va ben oltre il riconoscimento facciale in tempo reale, ma può consentire – con un utilizzo per l'appunto 'intelligente' dei dati acquisiti – di compiere analisi predittive, mettendo in stretta connessione le banche dati biometriche con i sistemi di tracciamento audio-video.

Sul punto, possiamo iniziare a parlare, in questo ambito come in altri ormai, di una videosorveglianza 'algoritmica' e 'automatizzata' con 'finalità predittive'.

Alla luce di ciò, mi chiedo se l'attuale impianto normativo (a partire dalla dimensione costituzionale e, via via, discendendo) sarà in grado di reggere questo vero e proprio volo pindarico indotto dalle nuove tecnologie e che rischia, in modo più o meno evidente, di erodere gradualmente l'esercizio pieno delle libertà costituzionali in nome del progresso e dell'esaltazione di una visione ossessivamente securitaria della dimensione sociale individuale e collettiva.

## Bibliografia

BARTOLE S., DE SENA P., ZAGREBELSKY V., Commentario breve alla Convenzione europea dei diritti dell'uomo e delle libertà fondamentali, Padova, CEDAM.

DE VERGOTTINI G., La difficile convivenza fra libertà e sicurezza: la risposta delle democrazie al terrorismo. Gli ordinamenti nazionali, in Atti del XVIII Convegno Annuale dell'Associazione (Bari, 17-18 ottobre 2003), Padova CEDAM, 2007.

DOGLIANI M., Il volto costituzionale della sicurezza, in Astrid Rassegna, n. 22/2010.

DOLCINI E., Un Paese meno sicuro per effetto del decreto-legge sicurezza, in Sistema Penale, 15 maggio 2025.

F. Q., Appello contro il decreto Sicurezza, il testo integrale e i nomi dei firmatari: da Zagrebelsky a De Siervo, in Il Fatto Quotidiano, 27 aprile 2025; Sul "pacchetto sicurezza" varato con decreto-legge, 9 aprile 2025, 1 (documento dell'Associazione italiana dei professori di diritto penale - AIPDP) disponibile sul portale [www.questonegiustizia.it](http://www.questonegiustizia.it)

GIUPPONI T. F., La sicurezza e le sue "dimensioni" costituzionali, in Forum di Quaderni Costituzionali, 2008.

GROSSO E., È compito della Repubblica. Note sul DDL Sicurezza, in Giustizia Insieme, 4 novembre 2024.

GROSSO G., Sicurezza: «Droni per sorvegliare la città». L'idea contro vandali e ladri, articolo del 30 aprile 2024 disponibile all'indirizzo [www.torinocronaca.it](http://www.torinocronaca.it).

MANETTI M., BORRELLO R. (a cura di), Videosorveglianza e privacy, Firenze, Angelo Pontecorbo Editore, 2010.

PACE A., Libertà e sicurezza. Cinquant'anni dopo, in Diritto e Società, n. 2/2013.

RUOTOLI M., Su alcune criticità costituzionali del c.d. pacchetto sicurezza (A.S. 1236), in Sistema Penale, 9 ottobre 2024.

Videosorveglianza (voce), in L. BOLOGNINI, E. PELINO (diretto da), Codice della disciplina privacy, Milano, Giuffrè Francis Lefebvre, 2024.

## **Diritto di accesso ad Internet e prospettive d'uso del Metaverso in Sanità – di Lorenzo Maria Lucarelli Tonini, Assegnista di Ricerca Università degli Studi Roma Tre, Dipartimento di Giurisprudenza**

*Il diritto di accesso a Internet non è più soltanto una questione tecnica, ma una condizione abilitante per esercitare diritti e adempiere doveri in un “non luogo” come il cyberspazio. La sua evoluzione, fino al metaverso, mostra effetti concreti anche su ambiti sensibili come la sanità e la protezione dei dati personali*

L'accesso a Internet, oggi configurato nella veste di diritto<sup>178</sup>, può essere considerato come una “nuova” situazione giuridica abilitatrice di ulteriori condizioni di vita che, tramite l'accesso ad un luogo, anzi un “non luogo”, ossia il cyberspazio, rende possibile esercitare diritti e adempiere doveri, originati dal mondo materiale oppure totalmente nuovi e sorti in ragione delle peculiarità di Internet stesso.

Tecnicamente, invece, per diritto di accesso ad Internet è da intendersi la materiale possibilità di disporre di un collegamento elettronico, tramite una connessione stabile e tecnologicamente adeguata, che consenta l'effettiva navigazione web. Oggi, non disporre di una rete Internet, oppure disporre di una rete non efficiente, ha infatti dirette ripercussioni sulla possibilità di esercitare i propri diritti e doveri, molti dei quali, per essere resi veramente effettivi, necessitano ormai delle reti.

Come testimonia nei tempi più recenti anche lo sviluppo del c.d. “metaverso”, si consideri come Internet abbia subito negli ultimi anni una grande evoluzione: se, inizialmente, l'uso di Internet era riservato esclusivamente al mondo professionale, l'odierno utilizzo, vista la facilità di accesso, riguarda nella stessa misura anche l'uso personale.

Inoltre, rispetto ad un interesse iniziale di carattere principalmente tecnologico, l'espansione di Internet ha oggi portato a ritenere come questa infrastruttura abbia invece sempre più valore di bene costituzionalmente protetto, il cui riferimento normativo può essere rinvenuto nell'art. 2 Cost, quale fattispecie aperta cui è possibile ricondurre la fonte di riconoscimento di diritti inviolabili non espressamente previsti, nonché nell'art. 3 Cost, quale addentellato normativo che riconosce il diritto all'uguaglianza formale e sostanziale, in questo caso richiamato dalla necessità che l'uso di Internet, e delle tecnologie in generale, sia aperto a tutti, allo stesso modo e non sia origine di possibili discriminazioni ma, invece, sia strumento di uguaglianza.

A riguardo, costante è stato il dibattito sulla qualificazione del diritto di accesso ad Internet quale diritto costituzionalmente protetto, con particolare riferimento alla necessità di un

---

<sup>178</sup> Il percorso di qualificazione dell'accesso ad Internet quale diritto è stato sicuramente segnato dal fondamentale contributo fornito dal Prof. Stefano Rodotà, il quale ne ha indirizzato i successivi sviluppi. A riguardo v. S. RODOTÀ, *Una costituzione per internet?*, in *Pol. dir.*, 3/2010, 337 ss.

espresso riconoscimento, oppure se i fondamenti costituzionali fossero già contemplati implicitamente da disposizioni, anche costituzionali, esistenti.

Seguendo la via del riconoscimento espresso, intesa a rafforzare e garantire l'inclusione della persona in un mondo sempre più digitale, tale così da sottrarre la possibilità di negare o limitare il diritto di accesso ad Internet da parte di terzi, l'accesso ad Internet è stato inizialmente studiato soprattutto in connessione con la libertà di espressione e il diritto di informazione, cui ha fatto seguito un'assimilazione dello stesso quale, appunto, libertà. Emblematici, in questo senso, sono i disegni di legge volti alla previsione di un art. 21bis della Costituzione, di cui è principale riferimento quello presentato nel 2010 dal Prof. Stefano Rodotà<sup>179</sup>, che così disponeva: «*Tutti hanno eguale diritto di accedere alla rete Internet, in condizione di parità, con modalità tecnologicamente adeguate e che rimuovano ogni ostacolo di ordine economico e sociale*».

Tuttavia, pur riconoscendo il merito di questo genere di proposta, con il passaggio dal web 1.0 al web 2.0 e oggi, con il metaverso, al web 3.0, seguito dal parallelo mutamento sociale, chiaro è come la qualificazione del diritto di accesso ad Internet nel limitato contesto informativo risulti essere ormai riduttiva, in quanto Internet può e deve essere considerato in relazione a molteplici diritti (e doveri) che oggi lo presuppongono, rappresentando ormai un vero e proprio luogo ove la persona agisce e si rappresenta.

Ne è seguita, invece, una qualificazione evolutiva, sicuramente più attuale e oggi punto di convergenza della dottrina maggioritaria, tesa a considerare il diritto di accesso ad Internet come un vero e proprio diritto sociale<sup>180</sup> che, come tale, per essere reso effettivo richiede l'intervento di un ulteriore soggetto, quale ad esempio lo Stato.

La proposta<sup>181</sup>, che ad oggi non ha avuto ancora ulteriore seguito formale, prevede che tutti abbiano «eguale diritto di accedere alla rete Internet in condizione di parità e con modalità tecnologicamente adeguate», demandando alla Repubblica il dovere di promozione delle «condizioni che rendono effettivo l'accesso alla rete Internet come luogo dove si svolge la personalità umana, si esercitano i diritti e si adempiono i doveri» e determinandone la limitazione in base a garanzie stabilite dalla legge e solo per atto motivato dall'autorità giudiziaria.

---

<sup>179</sup> S. RODOTÀ, *Audizione Comm. Affari Costituzionali*, Senato della Repubblica, 15 marzo 2015, seguita da Atti Parlamentari – Senato della Repubblica, n. 2485.

A tale configurazione sono seguite, sino ad oggi, nuove riproposizioni del diritto di accesso ad Internet incardinato nel contesto dell'art. 21bis Cost: ad es. Atti Parlamentari – Camera dei Deputati, n. 850 del 20 aprile 2013 e n. 327 del 13 ottobre 2022.

<sup>180</sup> G. D'IPPOLITO, *Il diritto di accesso ad Internet in Italia: dal 21 (-bis) al 34-bis*, in *MediaLaws*, 1/2021, 90; M. MENSI - P. ALLETTA, *Il diritto del web. Casi e materiali*, Padova, 2015, 40; P. PASSAGLIA, *Internet nella Costituzione Italiana: considerazioni introduttive*, in *Consulta online*, 2013, 22.

Per un'attenta disamina della possibilità di qualificare il diritto di accesso ad Internet come diritto sociale, v. P. TANZARELLA, *Accesso a Internet: verso un nuovo diritto sociale?*, in *Gruppo di Pisa*, 3 settembre 2012; G. DE MINICO, *Accesso a Internet tra mercato e diritti sociali nell'ordinamento europeo e nazionale*, in *federalismi.it*, 4/2018, 127 ss.

<sup>181</sup> Atti Parlamentari – Camera dei Deputati, n. 1136, *Proposta di legge costituzionale, Introduzione dell'articolo 34-bis della Costituzione, in materia di riconoscimento del diritto sociale di accesso alla rete internet*, 4 settembre 2018.

In tale ottica, il diritto di accesso ad Internet ha trovato un tentativo di formalizzazione nella proposta di previsione dell'art. 34-bis Cost., teso a qualificare espressamente il diritto di accesso ad Internet come diritto sociale e quindi comportando una pretesa soggettiva di prestazione nei confronti dello Stato, così come avviene ad esempio in tema di diritto alla salute o all'istruzione.

Per completezza, infine, si rileva come, a seguito del mancato sviluppo dei disegni normativi richiamati, già da qualche anno - nel 2015 precisamente - è stata emanata dalla "Commissione per i diritti e i doveri in Internet", costituita presso la Camera dei deputati, la "Dichiarazione dei Diritti in Internet"<sup>182</sup> che, pur priva di un effettivo valore normativo, riconosce tra i principi fondanti il cyberspazio, all'art. 2, il «diritto di accesso» a Internet e secondo la quale «*ogni persona ha eguale diritto di accedere a Internet in condizioni di parità, con modalità tecnologicamente adeguate e aggiornate che rimuovano ogni ostacolo di ordine economico e sociale*».

Come anticipato, attualmente costituisce una diretta declinazione dello sviluppo di Internet, da cui derivano nuove ed ulteriori implicazioni in termini di tutela dei diritti, l'affermarsi del fenomeno del "metaverso"<sup>183</sup>. A riguardo, è possibile fare riferimento, in via del tutto generica e non esaustiva, in mancanza di una definizione formalizzata ed univoca e di una disciplina direttamente riferibile<sup>184</sup>, ad una realtà "virtuale" e spesso "aumentata", caratterizzata da una particolare connessione con la realtà materiale, alla quale partecipa una «amalgama» di tecnologie<sup>185</sup>, nella quale sia possibile svolgere, attraverso un proprio *avatar*, un "gemello digitale"<sup>186</sup>, nonché attraverso l'uso di strumenti quali visori e comandi, molte delle attività comunemente praticate nel mondo fisico, attraverso un'esperienza immersiva in grado di accentuare ancora di più la propria esistenza *onlife* e tale da realizzare quasi una «simbiosi tra l'uomo e la tecnologia», che tradisce un «cambiamento del corpo e della persona» ma anche «una modifica poliedrica dei suoi rapporti con il mondo esterno»<sup>187</sup>, in ragione dell'ormai rinnovato «modo in cui la nostra mente, collegata al cervello, intende la materia»<sup>188</sup>.

Quanto detto, trova tuttavia la propria radice comune, consentendone il funzionamento e l'accessibilità, pur sempre nella possibilità di accedere ad Internet che, tuttavia, non è da solo un presupposto sufficiente per poter parlare di metaverso, il quale ne rappresenta, invece, una

---

<sup>182</sup> Camera dei Deputati - Commissione di studio sui diritti e i doveri relativi ad Internet, *Dichiarazione dei Diritti in Internet*, 20 luglio 2015.

<sup>183</sup> Come noto, infatti, il termine "metaverso" è stato utilizzato per la prima volta nel 1992, all'interno del romanzo "Snow Crash" dell'autore Neal Stephenson, ambientato in America in una realtà parallela e tecnologica.

<sup>184</sup> Sul punto v. G. LOFARO, *Applicazioni del metaverso nei servizi sanitari: governance della realtà estesa e blockchain per l'automazione dei procedimenti amministrativi*, in *Rivista di diritto amministrativo*, 4/2023, 1722; A. CELOTTO, *Il "metaverso" e delle sue implicazioni per l'ordinamento giuridico, audizione presso la Commissione Affari Costituzionali del Senato della Repubblica*, in senato.it, 6 luglio 2022.

<sup>185</sup> Tra le molteplici definizioni esistenti, sulle quali non vi è ancora oggi una effettiva convergenza, v. H. ULLAH et al., *Exploring the potential of metaverse technology in healthcare: applications, Challenges, and future directions*, in *IEEE Access*, 11, 2023.

<sup>186</sup> Approfondisce gli aspetti relativi al rapporto di specularità tra persona fisica e il proprio *avatar*, A. IANNUZZI, *Metaverso, Digital Twins e diritti fondamentali*, in *Inf. e dir.*, 2/2024, 41 ss.

<sup>187</sup> L. DI MAJO, *L'art. 2 della Costituzione e il "Metaverso"*, in *MediaLaws*, 2/2023, 37.

<sup>188</sup> G. CERRINA FERONI, *Il metaverso tra problemi epistemologici, etici e giuridici*, in *MediaLaws*, 1/2023, 22.

delle più recenti evoluzioni. Addirittura, è stato notato come, seppure Internet sia alla base di molte tecnologie, così come del metaverso, in questo caso la particolare complessità e fisionomia del fenomeno ne rappresentano una avanzatissima evoluzione, in continuo svolgimento, tale addirittura da «trascenderlo»<sup>189</sup>.

### Accesso ad Internet e tutela della salute

L'accesso ad Internet, come lo si consideri, è quindi direttamente connesso, anche se non esclusivamente delimitato, all'esercizio di ulteriori diritti, alcuni dei quali strettamente serventi la tutela della dignità della persona, consentendo di prescindere dalle condizioni personali, sociali, economiche o dal luogo ove questa si trovi e quindi esercitare un "esercizio diffuso" dei diritti. Ne costituisce un esempio assai evidente la tutela della salute, rispetto al quale Internet rappresenta ormai uno strumento necessario per garantire il corretto funzionamento dei sistemi sanitari più avanzati, sia per la stretta gestione amministrativa, sia per l'erogazione delle vere e proprie cure: in tale contesto viene, tuttavia, in rilievo un ulteriore diritto di carattere fondamentale, ossia il diritto alla protezione dei dati personali<sup>190</sup>, i quali alimentano o vengono a loro volta prodotti dalle tecnologie stesse usate in ambito clinico.

Tra i sistemi informatici più recenti, strettamente dipendenti da Internet e impiegati in ambito sanitario è ad esempio possibile fare riferimento al Fascicolo Sanitario Elettronico (FSE) e al Dossier sanitario elettronico (DSE), alla Cartella clinica elettronica, ai sistemi di comunicazione interna ed esterna (email o apposite piattaforme). È ancora possibile fare riferimento a tutti quei servizi che consentono di disporre online dei referti elettronici; alle app mediche; ai "medical devices" e ai dispositivi impiantabili, alla telemedicina e alle piattaforme di monitoraggio remoto; ai sistemi utilizzati per l'archiviazione dei dati e per adempiere al debito informativo verso le autorità; ai sistemi utilizzati per il trasferimento di dati tra dispositivi o strutture differenti, anche all'estero; ai sistemi utilizzati per l'analisi, il confronto e l'aggregazione dei dati sanitari, così come quelli utilizzati nella gestione dell'iter di fornitura e assistenza di strumenti e dispositivi medici stessi. Sempre più, infine, sta irrompendo nel settore sanitario, in maniera trasversale rispetto alle tecnologie suddette, l'utilizzo di sistemi virtualizzati (c.d. in *cloud*), dell'Intelligenza Artificiale, della realtà aumentata e della *blockchain*.

L'utilizzo di tutte queste tecnologie, ognuna delle quali pone differenti questioni rispetto alla tutela della persona e dei dati sanitari, anche di carattere problematico, comporta un rilevante trattamento di informazioni personali riferite ai pazienti, nel costante obiettivo di garantire una presa in carico e una tutela della persona sempre più efficiente e calata sulle necessità del

---

<sup>189</sup> Così R. BIFULCO, *Riverberi costituzionali del Metaverso*, in *MediaLaws*, 3/2023, 21.

<sup>190</sup> Relativamente al tema del trattamento dei dati personali relativi alla salute v. C. COLAPIETRO - F. LAVIOLA, *I trattamenti di dati in ambito sanitario*, in *Dirittifondamentali.it*, 2/2019.

singolo, quindi “personalizzata”, apprendo quindi ad una nuova frontiera del diritto alla salute che, così come disegnato dal legislatore costituente, è in continua evoluzione.

In tale contesto, la possibilità di trasferire e trattare dati con sempre maggiore facilità e intensità pone, inoltre, la questione relativa al luogo e alle modalità di archiviazione, in contemporamento con l'esigenza della loro protezione ma anche della necessaria circolazione: a riguardo, è sempre più concreta la realizzazione<sup>191</sup> da parte delle istituzioni europee<sup>192</sup> di uno spazio europeo per la condivisione dei dati sanitari, il c.d. “European Health Data Space” (EHDS)<sup>193</sup>, quale centralizzazione dei dati sanitari per una presa in carico condivisa, nel quale, appunto, confluiscano, tramite rete Internet, determinate categorie di dati provenienti dai sistemi informatici impiegati dalle singole strutture.

### **Una postilla sull'uso del Metaverso nel settore sanitario**

Quanto illustrato sino ad ora consente di anticipare come le tecnologie oggi disponibili nel settore sanitario siano numerose e molto diverse tra di loro, seppure spesso accomunate dall'uso di Internet quale infrastruttura di base. È possibile anche affermare che molte di queste tecnologie consentono, facilitandola, un'interazione, principalmente a distanza, tra i soggetti coinvolti, quali il paziente e il medico o la struttura sanitaria, così garantendo la tutela di un diritto fondamentale anche nei confronti di persone che non potrebbero spostarsi, sia per motivi legati a condizioni di salute sia per altri motivi, fornendo quindi un luogo che non sia un'alternativa di compromesso rispetto alla presenza fisica, quanto, invece, effettivamente sostitutivo e, possibilmente, migliorativo rispetto a quest'ultimo<sup>194</sup>.

Quanto al metaverso e alle potenzialità che questo può riservare, a fronte dell'iniziale interesse sorto e in seguito ad un successivo parziale affievolimento<sup>195</sup>, dovuto ad un insieme di concuse di carattere sia tecnologico (in particolare la complessità e disponibilità di questa tecnologia), che giuridico (principalmente riconducibili alla questione della certezza del diritto) ed economico<sup>196</sup>, occorre invece provare a domandarsi, a fronte di quelli che possono essere sicuramente alcuni problemi, se e come il metaverso possa aprire «una nuova frontiera nella sanità anche in riferimento all'evoluzione della medicina e dell'assistenza sanitaria verso il concetto di personalizzazione e prossimità», ove «i pazienti ed i medici possono essere ospitati

<sup>191</sup> Tale ecosistema è stato istituito con l'approvazione del “Regolamento (UE) 2025/327 del Parlamento europeo e del Consiglio, dell'11 febbraio 2025, sullo spazio europeo dei dati sanitari e che modifica la direttiva 2011/24/UE e il regolamento (UE) 2024/2847”.

<sup>192</sup> Sottolinea tale aspetto G. LOFARO, *Dati sanitari e e-Health europea: tra trattamento dei dati personali e decisione amministrativa algoritmica*, in *MediaLaws*, 2/2022, 200 ss.

<sup>193</sup> S. CORSO, *Lo spazio europeo dei dati sanitari: la Commissione Europea presenta la proposta di regolamento*, in *federalismi.it - Osservatorio di Diritto Sanitario*, 10 agosto 2022.

<sup>194</sup> Così G. CERRINA FERONI, *Il metaverso tra problemi epistemologici, etici e giuridici*, cit., 25.

<sup>195</sup> In questo senso, A. MALASCHINI, *Metaverso e diritto: su alcune questioni giuridiche della realtà virtuale*, in *Giustizia amministrativa*, 2/2023.

<sup>196</sup> Individua tali concuse A. IANNUZZI, *Metaverso, Digital Twins e diritti fondamentali*, cit., 36-37.

in nuovi spazi virtuali/fisici che hanno una disponibilità ed accessibilità continua»<sup>197</sup>, anche attraverso uno «sviluppo metaversico del biodiritto»<sup>198</sup>, che, grazie alla capacità «mimetica»<sup>199</sup> che lo connota consentirebbe, ad esempio, pur sempre nel rispetto della legge, di svolgere attività altrimenti non possibili nel concreto, perché complesse o rischiose. Del resto, già la telemedicina, pensata nelle diverse declinazioni della teleassistenza, della televista e del teleconsulto, è ormai utilizzata da molto tempo per fini in parte comuni. In tale contesto, ad un primo sguardo, l'uso del metaverso ne rappresenterebbe una importante evoluzione, in un'ottica di integrazione tra diverse e sempre nuove tecnologie.

Ciò che in questa sede si vuole evidenziare è, in particolare, come l'impiego del metaverso nel settore sanitario possa costituire un ulteriore sviluppo in tal senso ma, soprattutto, come ciò debba avvenire secondo condizioni certe e a valle di un bilanciamento tra la tutela della salute della persona, dei suoi dati personali e della sua dignità, quindi quale strumento atto a garantire la tutela di diritti fondamentali ma anche quale “luogo” ove questi, nel rispetto dell'equilibrio costituzionale, debbono essere ugualmente tutelati<sup>200</sup>.

Tale prospettiva va considerata e incentivata alla luce del fatto che molte delle tecnologie sin qui citate, tra cui anche la telemedicina, possono incontrare dei limiti derivanti dalla mancanza di una parvenza di “contatto” tra medico e paziente. Diverso, invece, sarebbe il caso in cui tale contatto avvenisse nel concreto ma a distanza, tramite un luogo ed un'identità virtuale, rendendo quindi possibile svolgere alcune delle attività che tipicamente un medico ritiene necessarie per visitare un paziente e al contempo acquisendo parametri importanti (anche tramite dati biometrici), ad esempio di movimento, altrimenti difficili da trasferire e valutare con una semplice televisita. Una simile prospettiva, con le dovute cautele, consentirebbe dunque un'integrazione tra ciò che oggi costituisce la telemedicina, con i vantaggi ottenibili tramite dispositivi indossabili, in un luogo nuovo, ossia il metaverso. Ne conseguirebbe il raggiungimento di molteplici risultati, sia in termini di presa in carico del paziente, sia in termini logistici ed economici, considerata la diretta ricaduta derivante dalla sua mancata presenza fisica all'interno della struttura sanitaria, con la possibilità, inoltre, di garantire ambienti personalizzati, ad esempio realizzati per specifiche patologie - quali quelle psichiatriche -, o ambienti “user-friendly” - per esempio destinati ai minori -, nei quali il gioco e la visita possano unirsi, rendendo più agevole l'attività dei medici e migliore l'esperienza del paziente.

Al contempo, si veda come l'utilizzo del metaverso possa avere positive ricadute anche ai fini della formazione sanitaria, consentendo ai discenti di confrontarsi in maniera più semplice con situazioni altrimenti difficili da replicare o controllare, quindi attraverso una formazione pienamente immersiva, anche di carattere pratico, ma esente dai rischi tipici.

---

<sup>197</sup> G. LOFARO, *Applicazioni del metaverso nei servizi sanitari: governance della realtà estesa e blockchain per l'automazione dei procedimenti amministrativi*, cit., 1724.

<sup>198</sup> A. IANNUZZI, *Metaverso, Digital Twins e diritti fondamentali*, cit., 42.

<sup>199</sup> R. BIFULCO, *Riverberi costituzionali del Metaverso*, cit., 44.

<sup>200</sup> Sottolinea tale aspetto L. DI MAJO, *L'art. 2 della Costituzione e il “Metaverso”*, cit., 37.

Quanto allo stato dell'arte, questo è fermo in Italia principalmente ad una fase di ricerca e primi sviluppi<sup>201</sup> di sistemi adeguati che consentano di conformarsi al meglio con le sfide poste da questa tecnologia: sia in termini di protezione dei dati personali, che di uso di sistemi di intelligenza artificiale, eventuale qualificazione dello strumento come dispositivo medico, nonché in termini di conformità con le normative medico-specialistiche.

Infatti, rispetto ad un ampio novero di previsioni di carattere sovranazionale di carattere generale<sup>202</sup> e poi anche nazionale, maggiormente specifiche<sup>203</sup>, volte a digitalizzare il settore sanitario, non si rinvengono invece provvedimenti specifici che regolino o prevedano (o vietino) l'utilizzo del metaverso nel settore sanitario; tuttavia, specifiche previsioni riguardano tecnologie quali la telemedicina o l'intelligenza artificiale, dalle quali si ritiene che, in mancanza di una normativa di settore, sia comunque possibile trarre, in prima battuta, un principio generale, consistente nell'interesse generale a rendere sempre più efficace la partecipazione sanitaria del paziente anche al di fuori delle strutture ospedaliere, per garantire la tutela della sua salute, come potrebbe avvenire con lo sviluppo e l'uso di sistemi "metaversici". A riguardo, poi, le più recenti normative in tema di protezione dei dati personali, intelligenza artificiale e cybersicurezza rimettono ad una valutazione e applicazione in concreto circa la sostenibilità e possibilità di sviluppare e utilizzare determinate tecnologie, a valle dell'applicazione di principi cardine che spesso si configurano invece come trasversali e non pensati per un singolo settore o strumento.

Qualche ulteriore elemento in tal senso, caratterizzato da maggiore specificità, è inoltre rinvenibile in una recente Raccomandazione europea<sup>204</sup>, con la quale il legislatore ha incoraggiato gli Stati ad introdurre strumenti tecnologici, in particolare di intelligenza artificiale, nonché basati su «tecnologie correlate», che facilitino la riabilitazione e reintegrazione delle persone sottoposte a regimi di restrizione di libertà, anche per facilitarne lo svolgimento delle attività di diagnosi e cura da remoto da parte dei medici. Leggendo tale documento, per certi

<sup>201</sup> Approfondisce il tema dell'uso del metaverso in sanità in altri paesi e nelle diverse branche della medicina G. LOFARO, *Applicazioni del metaverso nei servizi sanitari: governance della realtà estesa e blockchain per l'automazione dei procedimenti amministrativi*, in *Rivista di diritto amministrativo*, 4/2023, 1724.

<sup>202</sup> A titolo non esaustivo si richiamano la: Comunicazione COM(2004)356 - 30 aprile 2004; Comunicazione COM(2012) 736 final - 6 dicembre 2012; Comunicazione COM(2018) 233 final - 24 aprile 2018; Regolamento (Ue) 2021/522; Comunicazione COM(2024) 206 final - 22 maggio 2024; Comunicazione COM(2025) 10 final - 15 gennaio 2025.

<sup>203</sup> Si veda come in Italia gli interventi del legislatore, soprattutto in passato, siano stati maggiormente tesi all'istituzione e regolazione di singoli strumenti di sanità digitale e non alla creazione di un vero e proprio "ecosistema" della sanità digitale.

Solo negli ultimi anni questo processo di digitalizzazione sanitaria ha cambiato passo, adeguandosi ai riferimenti europei e acquisendo maggiore programmaticità, a partire dalla formalizzazione degli obiettivi di sanità digitale all'interno di documenti quali il "Piano triennale per l'informatica nella pubblica amministrazione", in base a quanto previsto dal Decreto legislativo 7 marzo 2005, n. 82 e dalla Legge 28 dicembre 2015, n. 208, art. 1, co. 513; dal "Patto per la Sanità Digitale", quale piano specifico di intesa tra Governo, Regioni e Province autonome del 24 giugno 2016 (previsto dalla Strategia per la crescita digitale 2014-2020 e dal "Patto per la Salute" 2014-2016) e dal "Piano Nazionale della Sanità Digitale 2022-2027". In tale contesto, centrale è il ruolo rivestito dal Piano Nazionale dei Ripresa e Resilienza (PNRR) – Mission 6 "Salute", che ha dedicato alla digitalizzazione della sanità, ed in particolare a specifici interventi, un'apposita quota parte dei fondi.

<sup>204</sup> Si vedano in particolare i punti 21 e 22 della Raccomandazione CM/Rec(2024)5 del Consiglio dell'Unione europea "regarding the ethical and organisational aspects of the use of artificial intelligence and related digital technologies by prison and probation services", del 9 ottobre 2024.

versi passato inosservato, sembra potersi affermare come il legislatore europeo stia ponendo le basi per lo sviluppo di ambienti “protetti” che facilitino, in determinati contesti, quali quello penitenziario - nello specifico - o sanitario - considerato in questo caso di riflesso -, il “contatto”. Di conseguenza, sembra potersi ritenere rientrante in tale previsione, anche in un’ottica evolutiva, costituendo una possibile tecnologia «correlata» eletta a tali scopi, proprio il metaverso.

Quanto agli aspetti squisitamente di carattere tecnico, un utile riferimento, nonché potenziale trampolino di lancio, può essere rinvenuto nella recente presentazione della Piattaforma Nazionale di Telemedicina<sup>205</sup> (PNT), quale piattaforma centralizzata che, seppure in maniera non esclusiva, rappresenterà l’infrastruttura di base per svolgere le attività tipiche della telemedicina. Rispetto a questo esempio, si auspica che anche lo sviluppo del metaverso nel settore sanitario possa seguire la medesima strada, alternativamente attraverso un’integrazione con la piattaforma suddetta, oppure attraverso lo sviluppo di una piattaforma, sempre di carattere pubblico e su scala nazionale<sup>206</sup>, appositamente ideata ed interoperabile, nella quale siano considerati, attraverso un procedimento di vaglio da parte delle istituzioni e delle autorità coinvolte, definibile di «costituzionalizzazione del Metaverso»<sup>207</sup>, anche in sinergia con potenziali soggetti privati (i providers), tutti gli aspetti normativamente richiamati.

## Bibliografia

BIFULCO R., *Riverberi costituzionali del Metaverso*, in *MediaLaws*, 3/2023.

CELOTTO A., *Il “metaverso” e delle sue implicazioni per l’ordinamento giuridico, audizione presso la Commissione Affari Costituzionali del Senato della Repubblica*, in *senato.it*, 6 luglio 2022.

CERRINA FERONI G., *Il metaverso tra problemi epistemologici, etici e giuridici*, in *MediaLaws*, 1/2023.

COLAPIETRO C. – LAVIOLA F., *I trattamenti di dati in ambito sanitario*, in *Dirittifondamentali.it*, 2/2019.

CORSO S., *Lo spazio europeo dei dati sanitari: la Commissione Europea presenta la proposta di regolamento*, in *federalismi.it - Osservatorio di Diritto Sanitario*, 10 agosto 2022.

D’IPPOLITO G., *Il diritto di accesso ad Internet in Italia: dal 21 (-bis) al 34-bis*, in *MediaLaws*, 1/2021.

DE MINICO G., *Accesso a Internet tra mercato e diritti sociali nell’ordinamento europeo e nazionale*, in *federalismi.it*, 4/2018.

<sup>205</sup> La Piattaforma è stata realizzata da Agenas quale soggetto attuatore per conto del Ministero della salute ed implementata nell’ambito del sub-investimento 1.2.3 “Telemedicina”, della Missione 6 Salute del PNRR (M6-C1).

<sup>206</sup> Quindi sottoposta ad una precisa *governance* di carattere statale, dove ci sia però spazio anche per l’apporto di carattere privato, quale solitamente principale sviluppatore dei diversi metaversi. Sottolinea tale aspetto G. CERRINA FERONI, *Il metaverso tra problemi epistemologici, etici e giuridici*, cit., 31.

<sup>207</sup> R. BIFULCO, *Riverberi costituzionali del Metaverso*, cit., 48.

- DI MAJO L., *L'art. 2 della Costituzione e il “Metaverso”*, in *MediaLaws*, 2/2023.
- IANNUZZI A., *Metaverso, Digital Twins e diritti fondamentali*, in *Inf. e dir.*, 2/2024.
- LOFARO G., *Applicazioni del metaverso nei servizi sanitari: governance della realtà estesa e blockchain per l'automazione dei procedimenti amministrativi*, in *Rivista di diritto amministrativo*, 4/2023.
- LOFARO G., *Applicazioni del metaverso nei servizi sanitari: governance della realtà estesa e blockchain per l'automazione dei procedimenti amministrativi*, in *Rivista di diritto amministrativo*, 4/2023.
- LOFARO G., *Dati sanitari e e-Health europea: tra trattamento dei dati personali e decisione amministrativa algoritmica*, in *MediaLaws*, 2/2022.
- MALASCHINI A., *Metaverso e diritto: su alcune questioni giuridiche della realtà virtuale*, in *Giustizia amministrativa*, 2/2023.
- MENSI M.- ALLETTA P., *Il diritto del web. Casi e materiali*, Padova, 2015.
- PASSAGLIA P., *Internet nella Costituzione Italiana: considerazioni introduttive*, in *Consulta online*, 2013.
- TANZARELLA P., *Accesso a Internet: verso un nuovo diritto sociale?*, in *Gruppo di Pisa*, 3 settembre 2012;
- RODOTÀ S., *Audizione Comm. Affari Costituzionali*, Senato della Repubblica, 15 marzo 2015, seguita da Atti Parlamentari – Senato della Repubblica, n. 2485.
- RODOTÀ S., *Una costituzione per internet?*, in *Pol. dir.*, 3/2010.
- ULLAH H. et al., *Exploring the potential of metaverse technology in healthcare: applications, Challenges, and future directions*, in *IEEE Access*, 11, 2023.

## **La rete europea dei Garanti per la protezione dei dati: struttura, funzioni e criticità del sistema di governance del GDPR – di Veronica Montozzi, dottoranda di Ricerca in Discipline Giuridiche Pubblistiche, Internazionalistiche ed Europee presso il Dipartimento di Giurisprudenza dell’Università degli Studi di Roma Tre**

*Il contributo analizza la rete europea dei Garanti per la protezione dei dati personali istituita dal GDPR con l’obiettivo di assicurare una tutela uniforme e coerente dei diritti fondamentali relativi al trattamento dei dati personali nello spazio giuridico dell’Unione europea.*

Dinanzi alla «assenza di una visione globale di privacy e della sua tutela universalmente condivisa» , l’Unione europea si è posta l’obiettivo di creare uno spazio dove garantire a pieno i diritti della persona rispetto a chiunque ne gestisca, tramite i dati, l’identità . A tal fine, recependo e razionalizzando l’esperienza maturata in Europa negli ultimi vent’anni , il Regolamento europeo n. 2016/679 sulla protezione dei dati (d’ora in poi GDPR) si caratterizza per l’intento di perfezionare l’assetto amministrativo che assicuri livelli uniformi di tutela del diritto alla protezione dei dati nello spazio giuridico europeo , istituendo una serie di organismi, che nel loro insieme formano una rete, la c.d. rete europea dei Garanti dei dati personali, denominazione con la quale si fa riferimento all’insieme di varie istituzioni che cooperano tra loro al fine di assicurare un’effettiva e coerente applicazione della disciplina contenuta nel GDPR.

I motivi che hanno guidato il legislatore europeo nel delineare il nuovo sistema di governance, basato sul collegamento tra diversi organismi, sono essenzialmente due. Il primo motivo, di carattere formale, concerne la tipologia dello strumento normativo scelto per regolare la materia della privacy: in luogo di una direttiva (la c.d. Direttiva “madre” 95/46/CE), infatti, viene adottato un Regolamento, ossia un atto di portata generale, direttamente applicabile in tutti gli Stati membri dell’Unione europea, il quale, pur prevedendo diverse clausole di flessibilità , impone a tutti gli Stati membri l’applicazione omogenea di una medesima disciplina. Il GDPR, infatti, fermo restando la validità degli obiettivi e dei principi stabiliti dalla precedente direttiva 95/46, nasce dalla necessità di porre rimedio alla frammentazione della disciplina sulla protezione dei dati personali nell’Unione europea e dalla rilevazione della diffusa incertezza giuridica concernente l’applicazione della normativa .

L’adozione di una disciplina omogenea come quella offerta dal Regolamento della privacy è apparsa, quindi, funzionale a garantire in tutta l’Unione un elevato livello di tutela dei diritti e delle libertà fondamentali della persona fisica, con riguardo al trattamento dei dati personali. Il GDPR infatti comporta in primo luogo una medesima regolazione in tutto il territorio dell’Unione europea al fine di tutelare un diritto che, assumendo la veste di diritto fondamentale, sancito dall’art. 8 della Carta dei diritti fondamentali dell’Unione europea e

dall'art. 16 TFUE, richiede omogeneità di regolazione nei confronti di tutti coloro che risiedono nel suo territorio ; in secondo luogo, persegue l'obiettivo di incentivare la libera circolazione dei dati per il tramite di una ragionevole certezza delle modalità con le quali tale regolazione è in concreto applicata, creando uno "spazio fisico" comune entro cui sono dettate le medesime regole .

Il secondo motivo, di carattere sostanziale, attiene invece al contesto nel quale si sono inserite le diverse fonti . Rispetto dunque all'assetto antecedente la riforma del 2016 , caratterizzato dall'assenza di social network e motori di ricerca, aventi quale principale caratteristica quella di rendere il mondo interconnesso, nell'odierno mondo digitale si rende evidente la necessità di assicurare un'omogenea applicazione della disciplina sulla privacy su tutto il territorio dell'Unione europea e di creare un network di sorveglianza a presidio del diritto alla protezione dei dati e dell'omogeneità della sua tutela, essendo chiaro che l'esigenza di uniformità della normativa non possa prescindere da un enforcement in grado di assicurare l'effettiva coerenza da parte delle singole legislazioni nazionali . La dematerializzazione degli strumenti di circolazione delle informazioni e l'assenza di confini geografici del web impongono, infatti, la concertazione di decisioni delle autorità per assicurare un'efficace attuazione del GDPR e «realizzare un clima di fiducia del pubblico che consentirà lo sviluppo dell'economia digitale in tutto il mercato interno» .

La rete europea è stata quindi elaborata dal legislatore europeo al quale è apparsa evidente, fin da subito, l'esigenza di un meccanismo di armonizzazione che, da un lato, cogliesse le diversità inevitabili tra le posizioni delle Autorità dei paesi firmatari, le quali risultavano assai differenti tra di loro a causa delle normative a loro volta differenti le une dalle altre, e dall'altro, soddisfacesse l'esigenza di una necessaria applicazione coerente della disciplina.

Sebbene l'applicazione del GDPR abbia rappresentato un successo sin dai primi anni della sua entrata in vigore, avendo rafforzato la protezione dei dati come diritto fondamentale e armonizzato l'interpretazione dei principi in materia , grazie anche – e soprattutto – al nuovo sistema di governance istituito, di recente è stato rilevato l'insufficiente livello di cooperazione tra le Autorità dovuto alla frammentazione delle procedure e delle prassi nazionali, che determina «incertezza del diritto e aumenta i costi per le imprese (ad esempio a causa della necessità di produrre documenti differenti nei vari Stati membri), perturbando la libera circolazione dei dati personali nell'UE, pregiudicando le attività commerciali transfrontaliere e ostacolando la ricerca e l'innovazione in relazione a sfide sociali urgenti» .

Nei paragrafi che seguono, dunque, si cercherà di illustrare il complesso funzionamento della rete europea dei Garanti, partendo da un inquadramento dei soggetti che ne fanno parte e dei diversi strumenti che ne assicurano la cooperazione e il dialogo, analizzando, infine, quali sono le criticità che attualmente impediscono l'efficace funzionamento del sistema predisposto dal legislatore europeo nel GDPR e le prospettive di riforma.

## **Autorità di controllo indipendenti e Autorità di controllo “capofila”**

Nella realizzazione di un sistema europeo prodromico ad un’omogenea tutela dei dati personali, un ruolo fondamentale è coperto dalle autorità indipendenti nazionali, in quanto, da una parte, sorvegliano l’attuazione e la corretta implementazione delle regole sovranazionali nei rispettivi territori, dall’altra, garantiscono, attraverso meccanismi di coerenza e di stretta cooperazione, standard uniformi di tutela in tutto il territorio europeo .

L’art. 51 del Regolamento dispone l’istituzione di una o più autorità pubbliche di controllo indipendenti in ciascun Stato membro, aventi il compito di sorvegliare l’applicazione del Regolamento al fine di tutelare i diritti e le libertà fondamentali delle persone fisiche con riguardo al trattamento, e di agevolare la libera circolazione dei dati personali all’interno dell’Unione.

Occorre osservare che le Autorità di controllo in tema di privacy avevano già fatto il loro ingresso nella normativa europea con la Direttiva 95/46/CE, la quale, all’art. 28, disponeva che ciascuno Stato membro dovesse dotarsi di un soggetto pubblico, incaricato di sorvegliare, nel proprio territorio, l’applicazione delle disposizioni di attuazione della Direttiva. Nella Direttiva in esame, le Autorità erano investite soprattutto di poteri consultivi, svolgendo però, fin dalla loro comparsa, anche il compito di sviluppare, all’interno degli Stati membri, una «costante attività di implementazione della cultura della protezione dei dati personali» .

Il ruolo delle Autorità indipendenti è stato implicitamente modificato nella transizione dalla direttiva 95/46/CE al regolamento (UE) 2016/679, poiché il paradigma introdotto da quest’ultimo statuisce un rapporto diretto tra Autorità indipendenti di controllo e Regolamento, esigendo che siano immediatamente applicabili le funzioni di sorveglianza e controllo sulle disposizioni ivi contenute. Ciò ha determinato la centralità del ruolo dei Garanti nazionali, nel dare piena e coerente attuazione alle previsioni del regolamento e delle conseguenti norme interne di adeguamento, avendo peraltro assunto la capacità di adottare decisioni vincolanti anche al di fuori del territorio statale in cui operano.

In questa pianificazione, il legislatore europeo lascia agli Stati membri margini di manovra per disciplinare l’adozione delle previsioni di adeguamento al Regolamento in modo da renderle compatibili con l’ordinamento interno. Conseguentemente, dunque, ogni Autorità è competente anche per l’applicazione della legislazione nazionale nelle materie in cui il GDPR consente al legislatore dello Stato membro di legiferare.

Al fine di eseguire il compito fondamentale di “sorvegliare” la corretta applicazione del Regolamento, l’art. 58 investe le Autorità di controllo di una serie di poteri che possono distinguersi in tre categorie: poteri di indagine (art. 58, par. 1), poteri correttivi (art. 58, par. 2) e poteri consultivi (art. 58, par. 3). A tali poteri si affianca la facoltà, statuita con l’art. 83 del Regolamento, di infliggere sanzioni amministrative pecuniarie, individuate sulla base di criteri di effettività, proporzionalità e dissuasività .

In riferimento all’ambito territoriale di competenza di ciascuna Autorità di controllo , l’art. 55 del Regolamento stabilisce che esse esercitano i propri poteri ed eseguono i propri compiti nel territorio dello Stato cui appartengono, mentre il Considerando n. 122 precisa i confini della giurisdizione avendo riguardo al trattamento esaminato.

La compresenza sul territorio dell’Unione di una o più Autorità di controllo per ciascun Stato membro, nonché la possibilità per le singole autorità di essere investite di poteri ulteriori, determinano tuttavia il rischio di applicazioni difformi della normativa nei diversi Stati, soprattutto nelle fattispecie in cui le operazioni di trattamento coinvolgono una pluralità di Paesi membri, ossia in caso di trattamento transfrontaliero dei dati personali . In queste due diverse tipologie di casi, si rende evidente il rischio che una pluralità di Autorità nazionali diverse si pronunci, in modo anche comprensibilmente difforme, sulla conformità al GDPR di trattamenti di dati operati in più stabilimenti in diversi Paesi o che, pur effettuati da un unico stabilimento situato in un solo Paese, riguardino dati di interessati appartenenti ad altri Stati.

In considerazione quindi dell’esigenza di uniformità dell’applicazione della disciplina, è apparso ragionevole cercare di evitare che trattamenti di dati svoltisi sul territorio di uno Stato, ma aventi effetti anche in altri, possano essere sottoposti al sindacato di diverse Autorità di controllo – ciascuna competente per gli effetti provocati sul proprio territorio – prevedendo invece che, in caso di trattamenti che comportino la circolazione transfrontaliera di dati, la competenza a controllare i trattamenti stessi sia affidata a un’unica Autorità competente. A tal riguardo, il Regolamento istituisce all’art. 56 un altro componente della rete europea, la c.d. Autorità di controllo “capofila” (Lead Supervisory Authority), ossia l’autorità competente nel caso di trattamento di dati personali a carattere transfrontaliero, in deroga al criterio di imputazione territoriale della competenza di cui al par. 1 dell’art. 55, la quale assume in tali ipotesi un ruolo chiave essendo chiamata a coordinare la cooperazione con le altre Autorità appartenenti agli Stati membri interessati in modo da assicurare coerenza nell’applicazione del Regolamento, certezza del diritto e omogeneità nella sua attuazione anche laddove siano coinvolti più Stati .

Nell’individuazione dell’Autorità capofila fra i diversi Garanti nazionali interessati, a ben vedere, occorre tenere distinte le due fattispecie di trattamento transfrontaliero. Allorché infatti la qualificazione del trattamento come transfrontaliero discenda dalla circostanza che esso produca effetti che coinvolgono interessati che si trovano in diversi Paesi membri, l’autorità capofila viene individuata in quella dello Stato in cui è collocato l’unico stabilimento del titolare o del responsabile. Qualora, invece, il trattamento dei dati assuma la qualifica di transfrontaliero in quanto vi sia una molteplicità di stabilimenti collocati in più Stati dell’Unione, l’Autorità capofila è identificata in quella competente in relazione al luogo in cui è situato lo stabilimento principale del soggetto attivo del trattamento, inteso come lo stabilimento che effettivamente prende le decisioni in ordine a finalità e mezzi del medesimo, avendone per statuto il potere . Spetta, poi, al titolare/responsabile l’onere di provare che lo stabilimento, individuato come “principale”, svolga un «esercizio reale ed effettivo di attività gestionali o decisionali rispetto al trattamento di dati personali», ancorando la prova a criteri

oggettivi e a elementi probatori dimostrabili, non potendosi invece fondare esclusivamente su dichiarazioni rese dalla società .

In caso di opposizione di un Garante che reclami tale ruolo, l'individuazione della capofila spetta in ultima istanza al Comitato europeo.

Possono esserci casi “borderline” in cui è difficile stabilire quale sia lo stabilimento principale o determinare dove siano assunte realmente le decisioni inerenti al trattamento dei dati, come ad esempio l'ipotesi in cui un'impresa abbia più stabilimenti in Europa, nessuno dei quali sia la sede dell'amministrazione centrale e in nessuno di essi si prendano decisioni sul trattamento. In tali ipotesi, il GDPR non fornisce soluzioni chiare, prevedendo che spetti alla società indicare lo stabilimento che funge da stabilimento principale e, nel caso in cui non fosse possibile, sarà compito dell'Autorità di controllo del territorio interessato procedere all'individuazione dello stesso .

A fronte di difficoltà interpretative già nell'individuazione dell'Autorità Capofila, il testo del Regolamento adottato circonda questo istituto di molte garanzie e limitazioni, come dimostrano i complessi passaggi che compongono i meccanismi di cooperazione descritti negli artt. 56 e 60, nonché quelli relativi ai meccanismi di coerenza.

## **Il Comitato europeo per la protezione dei dati**

Accanto alle autorità nazionali, le esigenze di armonizzazione delle regole che governano la protezione dei dati personali hanno assunto una rilevanza tale da suggerire al legislatore l'opportunità di istituire un organismo dotato di compiti più ampi e pregnanti, quale il Comitato europeo per la protezione dei dati (di seguito, Comitato), o European Data Protection Board (EDPB). Esso è un organismo dell'Unione, dotato di personalità giuridica (art. 68, par. 1 GDPR), avente dunque la titolarità di tutte le situazioni giuridiche attive e passive e una responsabilità diretta per gli atti compiuti nell'esercizio delle sue funzioni, e si configura come un ente di diritto pubblico europeo, indipendente. Il Comitato sostituisce, senza soluzione di continuità, il precedente Gruppo di lavoro ex. Art. 29 (Working Party on Article 29), ossia l'organismo istituito dalla Direttiva che annoverava e coordinava la cooperazione tra le diverse Autorità indipendenti. Tuttavia, lungi dall'essere un mero successore, esso si differenzia in quanto dispone di maggiori poteri e della personalità giuridica, non prevista per il precedente organo , che gli consentono di ricoprire meglio il ruolo di “organo di chiusura del sistema” , il cui principale compito consiste nel vigilare sulla coerente applicazione delle norme del Regolamento, coordinando le diverse Autorità di controllo .

Come previsto ai sensi dell'art. 68, par. 3 – e specificato dal Considerando n. 139 – il Comitato è composto dalle figure apicali delle Autorità di controllo di ciascuno Stato membro (ad oggi 27), le quali partecipano con diritto di voto, nonché dal Garante europeo della protezione dei

dati o dai relativi rappresentanti, aventi diritti di voto specifici e limitati. Inoltre, partecipano alle attività del Comitato anche la Commissione europea e i rappresentanti delle autorità dell’Area Economica Europea (Islanda, Liechtenstein e Norvegia), avendo il diritto di esprimere le proprie posizioni su tutte le questioni discusse, ma restando privi del diritto di voto. Il Comitato adotta i propri atti a maggioranza, qualificata o semplice, rendendoli così reale espressione della volontà dell’istituzione e non dei singoli componenti, in modo tale che essi siano fortemente incentivati a rispettare l’indirizzo concordato.

In ragione di tale autorevole composizione, al Comitato viene riconosciuta, dunque, una funzione di nomofilachia, essendo stato istituito con la principale finalità di monitorare e assicurare la coerente applicazione del Regolamento (art. 70, par. 1). In questa prospettiva il GDPR gli assegna una serie di poteri e compiti, assolti dall’organo in piena indipendenza (art. 69), e tutti convergenti nell’indirizzare le pratiche delle singole Autorità di controllo di tutti gli Stati membri verso un’applicazione coerente delle norme sulla protezione dei dati, promuovendone al contempo la cooperazione . Il Comitato dunque esercita una costante attività di sorveglianza relativa all’applicazione corretta del Regolamento, non limitandosi all’esercizio di semplici funzioni consultive, come nel caso dei pareri emessi ai sensi dell’art. 64, ma espletando anche un potere decisionale, previsto dall’art. 65, in relazione alle controversie che possono sorgere tra le diverse Autorità di controllo. In quest’ottica, il Comitato detiene anche il potere di orientare le attività della Commissione – organo esecutivo e promotore del processo legislativo nell’Unione – nella direzione di un’omogenea applicazione del regolamento, svolgendo un’attività consultiva nell’ambito della quale fornisce ad essa indicazioni in merito a qualsiasi questione inerente alla protezione dei dati personali. Sempre con riguardo allo scopo di assicurare l’applicazione coerente del Regolamento, ai sensi dell’art. 70, par. 1, lett. e, il Comitato svolge l’ulteriore e fondamentale compito di elaborare orientamenti generali, esaminando, di propria iniziativa o su richiesta di uno dei suoi membri o della Commissione, questioni relative all’applicazione del Regolamento stesso, pubblicando, di conseguenza, linee guida, raccomandazioni e migliori pratiche per chiarire e promuovere una comprensione comune delle leggi sulla protezione dei dati dell’UE.

### **Il Garante europeo della protezione dei dati**

Tra i soggetti che compongono la rete europea dei Garanti, deve infine annoverarsi anche il Garante europeo della protezione dei dati (di seguito, GEPD), o anche European Data Protection Supervisor (EDPS), che opera quale autorità di sorveglianza indipendente, istituita con il Regolamento (CE) n. 45/2001, e nominato dal Parlamento e dal Consiglio dell’Unione europea per mandati rinnovabili di cinque anni. Esso svolge il primario compito di garantire che le istituzioni e gli organi dell’UE rispettino il diritto dei cittadini alla vita privata e alla protezione dei dati in sede di trattamento dei dati personali e di elaborazione di nuove politiche.

Attualmente, i doveri e i poteri del Garante europeo della protezione dei dati, nonché la sua l'indipendenza istituzionale in qualità di autorità di controllo, sono definiti nel regolamento (UE) 2018/1725, che ha abrogato il precedente regolamento e che stabilisce altresì le norme relative alla protezione dei dati nelle istituzioni dell'UE. Il GEPD assolve diversi compiti, tra cui in primis, quello di sorvegliare il trattamento dei dati personali da parte dell'amministrazione dell'UE allo scopo di assicurare il rispetto delle norme sulla privacy, in cooperazione con i responsabili della protezione dei dati presenti in ogni istituzione o organismo comunitario. Il Garante europeo collabora, inoltre, con le amministrazioni nazionali dei paesi dell'UE per assicurare la coerenza nell'ambito della protezione dei dati, ricoprendo il ruolo di membro del Comitato europeo con diritti di voto specifici e limitati all'adozione di decisioni che riguardano principi e norme applicabili ad istituzioni, organi, uffici e agenzie dell'Unione .

## **Cooperazione e coerenza**

Come anticipato, per far fronte all'inevitabile rischio di applicazioni difformi della normativa nei diversi Stati pericolo e assicurare l'uniforme applicazione della disciplina in materia di protezione dei dati personali, l'art. 51, par. 2, specificando quanto previsto nel Considerando n. 123, richiede che le autorità di controllo si impegnino a contribuire alla coerente applicazione del presente regolamento in tutta l'Unione, cooperando tra loro e con la Commissione, «così da tutelare le persone fisiche in relazione al trattamento dei loro dati personali e facilitare la libera circolazione di tali dati nel mercato interno». In vista di tale scopo nel Capo VII, rubricato, “Cooperazione e Coerenza”, il legislatore europeo disciplina i rapporti tra le Autorità per il tramite di due meccanismi: il meccanismo di cooperazione – finalizzato a garantire un necessario e continuo dialogo tra le Autorità di controllo interessate, al fine di individuare una soluzione congiunta che assicuri l'uniforme applicazione della disciplina – e il meccanismo di coerenza che, ponendosi al di sopra del primo , interviene in tutti i casi in cui, nonostante il dialogo instauratosi tra le Autorità di controllo capofila e le singole Autorità interessate, ovvero, tra le singole Autorità interessate, non si riesca a pervenire ad una decisione condivisa, coinvolgendo anche la Commissione europea e il Comitato.

## **Il meccanismo “one stop shop” e altre forme di cooperazione**

A differenza della direttiva 95/46/CE, che in un'unica disposizione (art. 28 direttiva 95/46/CE), imponeva alle autorità unicamente di scambiarsi «ogni informazione utile», il Regolamento (UE) 2016/679 dedica numerose disposizioni alla cooperazione al punto che non appare

improprio sussudere l'esistenza di un principio generale di leale collaborazione fra autorità di controllo ,

Quali nuove estrinsecazioni di tale principio, il Regolamento prevede precisi strumenti volti a garantire un concreto e costante dialogo tra le singole autorità di controllo. Essi sono il procedimento “one stop shop”, l'assistenza reciproca e le operazioni congiunte, che costituiscono un significativo passo in avanti nel processo di integrazione europea della Data protection.

Il meccanismo one stop shop, detto anche “meccanismo dello sportello unico”, è un meccanismo di cooperazione, stabilito dall'art. 60 GDPR, che disciplina i rapporti tra l'Autorità Capofila e le altre autorità di controllo. Esso si sostanzia in un meccanismo di co-decisione, dove vi è un'unica autorità con cui l'impresa si interfaccia, la c.d. Autorità capofila, la quale è competente a emanare nei suoi confronti la (unica) decisione finale, essendo però obbligata a interpellare tutte le Autorità interessate prima di assumere qualsiasi provvedimento che riguardi un titolare o responsabile, al fine di raggiungere un consenso condiviso in merito alla misura da adottare e garantire così la coerenza.

L'iter procedurale prende avvio su iniziativa di un interessato, il quale può proporre reclamo nello Stato dove risiede o dove lavora, denunciando la violazione di norme in materia di protezione dei dati personali da parte di un titolare del trattamento.

Ricevuto il reclamo e riscontrato il carattere transfrontaliero del trattamento, l'Autorità nazionale si rivolge all'Autorità capofila competente che, nel termine di tre settimane, deve decidere se trattare il caso (art. 56.4) secondo la procedura di cooperazione, oppure non trattare il caso (art. 56.5) rimettendolo alla competenza dell'Autorità di controllo nazionale (che ha ricevuto il reclamo).

Nel caso in cui la Capofila decida di attivare la procedura di cooperazione, è tenuta a condividere con le altre autorità interessate tutte le informazioni utili sulla questione e a trasmettergli un proprio progetto di decisione, valutando adeguatamente il parere da loro espresso su di esso.

Entro il termine di quattro settimane, ciascuna autorità di controllo può, infatti, sollevare una obiezione «pertinente e motivata» al progetto di decisione (art. 60, par. 4), ponendo l'autorità capofila innanzi alla seguente alternativa: se non reputa di dare seguito all'obiezione pertinente e motivata – oppure se ritiene che difettino questi due requisiti – essa deve sottoporre la questione al meccanismo di coerenza, di cui all'art. 63; qualora, invece, l'autorità capofila intenda dare seguito all'obiezione, è tenuta a modificare il progetto di decisione e a trasmetterlo nuovamente alle altre autorità di controllo interessate, così da ottenere il loro parere anche sulla versione riveduta.

L'adeguata valutazione dell'opinione delle altre autorità di controllo da parte dell'autorità capofila costituisce, infatti, un passaggio fondamentale del procedimento che conduce all'adozione di un provvedimento finale che sia realmente condiviso da tutte le Autorità . Tale valutazione non viene meno neanche nell'ipotesi in cui nessun Garante sollevi obiezioni al

progetto di decisione trasmesso dall'Autorità capofila, in quanto, in tal caso, si presume che tutte le autorità coinvolte siano d'accordo sul progetto, che diviene per loro vincolante in forza di un consenso implicito.

L'Autorità capofila adotta poi la decisione finale e la notifica allo stabilimento principale o allo stabilimento unico, informando le altre Autorità interessate, tra cui l'autorità nazionale che ha ricevuto il reclamo, che deve darne notizia al reclamante (art. 60.7).

La complessa articolazione del procedimento di cui all'art. 60 GDPR, è espressione della difficoltà di bilanciare le prerogative delle autorità di controllo nazionali con l'innovativo ruolo dell'Autorità capofila, in quanto la decisione di quest'ultima, di regola, si sostituisce a quelle che avrebbero potuto adottare – nell'ambito delle rispettive competenze territoriali – i diversi Garanti nazionali, venendo così assicurata l'omogeneità applicativa della disciplina della Data protection e, quindi, la certezza del diritto dell'Unione .

A questi fini, è previsto a carico del titolare o del responsabile, destinatario della decisione dell'autorità capofila, l'obbligo di garantire la conformità a essa di ogni trattamento realizzato in tutti i propri stabilimenti all'interno dell'UE (art. 60, par. 10). In tal modo, infatti, il meccanismo one stop shop consente l'estensione degli effetti della decisione in tutti i Paesi membri in cui operano i soggetti attivi che ne sono destinatari .

Il meccanismo appena descritto non si applica, tuttavia, nel caso in cui l'oggetto dei reclami o delle violazioni del regolamento riguardi unicamente uno stabilimento situato in uno Stato membro, oppure incida in modo sostanziale solo su interessati ivi ubicati, spettando così la competenza a decidere all'autorità garante dello Stato in cui si trovano lo stabilimento o gli interessati (art. 56, par. 2), fermo restando anche in tali fattispecie, l'obbligo di cooperazione fra autorità amministrative indipendenti, per il tramite di altri strumenti previsti dal Regolamento.

Il Considerando n. 128, inoltre, nega la competenza dell'Autorità Capofila, e dunque l'applicazione del meccanismo dello sportello unico, nei casi in cui il trattamento, oggetto del reclamo, sia effettuato da autorità pubbliche o da organismi privati nell'interesse pubblico, rimanendo quindi competente in queste ipotesi l'Autorità di controllo dello Stato in cui sono stabiliti l'autorità pubblica o l'organismo privato.

Occorre notare che, quella prevista nell'ambito del meccanismo one stop shop non è l'unica forma di cooperazione fra autorità di controllo disciplinata dal regolamento (UE) 2016/679, in quanto, animato dall'intento di garantire una omogenea applicazione della disciplina europea in tutti i Paesi membri, il legislatore europeo ha tipizzato ulteriori strumenti di collaborazione tra Garanti della privacy nazionali, quali l'assistenza reciproca e le operazioni congiunte .

L'assistenza reciproca, disciplinata ai sensi dell'art. 61 GDPR, si concretizza, in particolare, nelle richieste di informazioni e misure di controllo (quali le richieste di autorizzazioni e consultazioni preventive) e le richieste di effettuare ispezioni e indagini in un diverso Stato membro, da parte di un'Autorità alle altre. L'autorità destinataria delle richieste deve darvi seguito senza ingiustificato ritardo e comunque entro un mese dal ricevimento della richiesta

(par. 2). È da sottolineare la natura obbligatoria dell’assistenza reciproca, che la distingue dai generici inviti alla cooperazione contenuti, invece, nell’art. 28 par. 6 direttiva 95/46/CE. È pertanto previsto un esplicito divieto per le autorità destinatarie delle richieste di apporre un rifiuto, salvo due casi tassativamente descritti dalla norma al par. 4 . Al di fuori di queste ipotesi, qualora l’A. richiesta non trasmetta alcun riscontro nei termini previsti, l’Autorità di controllo richiedente può adottare “misure provvisorie”, analoghe a quelle di cui all’art. 66, par. 1 GDPR, destinate ad operare all’interno del proprio territorio.

Le operazioni congiunte delle autorità di controllo consistono invece in azioni comuni, incluse indagini e misure di contrasto, cui prendono fisicamente parte membri o personale appartenenti alle Autorità di controllo di altri Stati membri. A queste attività, in caso di trattamenti transfrontalieri di dati personali, hanno diritto di partecipare le Autorità degli Stati dell’UE in cui il titolare o il responsabile ha degli stabilimenti, oppure quelle dei Paesi membri in cui vi sia la probabilità che il trattamento abbia un «impatto negativo sostanziale» su un «numero significativo di interessati» ai sensi dell’art. 62 GDPR.

Entrambi questi strumenti possono essere utilizzati, tuttavia, anche al di fuori dei suddetti casi obbligatori, ognqualvolta un’autorità di controllo ritenga proficuo avvalersi del supporto di un’altra per azioni da condurre sul proprio territorio.

### **Il meccanismo di coerenza: gli strumenti nelle mani del Comitato europeo**

A tutela dell’uniformità di azione da parte delle diverse Autorità di controllo, riprendendo quanto espresso dal Considerando n. 135 , l’art. 63 sancisce che «al fine di contribuire all’applicazione coerente del presente regolamento in tutta l’Unione, le autorità di controllo cooperano tra loro e, se del caso, con la Commissione mediante il meccanismo di coerenza stabilito nella presente sezione». Qualora, infatti, le forme di cooperazione amministrativa fra Garanti nazionali non conducessero a una soluzione condivisa in merito alla decisione da adottare, l’ordinamento europeo prevede l’instaurazione di un dialogo tra le Autorità nazionali e la Commissione, dialogo che si svolge nell’ambito del Comitato europeo per la protezione dei dati personali che assume un ruolo centrale .

Il legislatore europeo ha dedicato un’apposita sezione del Regolamento al meccanismo di coerenza (Sezione II del Capitolo VII, rubricata “Coerenza”), intendendo, con il termine “coerenza”, tanto il meccanismo quanto l’obiettivo da raggiungere tramite esso, che si ricerca in fase di applicazione delle misure da parte dell’Autorità di controllo per la protezione dei dati.

Tale meccanismo è infatti composto da diversi strumenti, posti nelle mani del Comitato europeo, tramite i quali svolge la funzione di coordinare le attività delle autorità di controllo dei diversi Stati membri e della Commissione .

Il primo strumento, previsto dall’art. 64, si estrinseca nel potere del Comitato di adottare pareri che consentono di individuare una soluzione unitaria in merito a determinate questioni che possono presentarsi nel caso concreto. L’articolo si compone di due paragrafi: nel paragrafo n. 1 vengono contemplate le ipotesi tassative in cui il parere da parte del Comitato

deve essere obbligatoriamente richiesto. In particolare, esse concernono tutte le decisioni delle autorità di controllo rese nell’ambito di fattispecie in cui è consentito ai singoli Stati membri mantenere un ampio margine di manovra rispetto all’applicazione del Regolamento (elenco di trattamenti soggetti al requisito di una valutazione di impatto, conformità al Regolamento di codici di condotta, requisiti per accreditamento di organismi, clausole contrattuali, ecc.), rendendo evidente dunque che in questo caso l’intervento del Comitato sia diretto ad assicurare la coerenza delle decisioni assunte dalle diverse autorità nazionali. Nel paragrafo n. 2, invece, si attribuisce al Comitato un generico potere di adozione di pareri che possono essere discrezionalmente richiesti da parte di qualsiasi autorità di controllo, dalla Commissione europea o dal Presidente del Comitato, in merito a questioni di applicazione generale della normativa, ovvero questioni destinate a produrre effetti in più di uno Stato membro, come ad esempio nell’ipotesi in cui un’autorità di controllo competente non si conformi agli obblighi relativi all’assistenza reciproca di cui all’art. 61, o allo svolgimento di operazioni congiunte ai sensi dell’art. 62.

In entrambe le ipotesi previste dall’articolo, il Comitato può rifiutarsi di emettere il parere qualora rilevi di essersi già pronunciato in passato su una questione analoga (art. 64, par. 3).

Sempre alla luce del più generale obiettivo di garantire la coerente applicazione del Regolamento, l’art. 65 GDPR prevede un altro strumento, ossia il potere del Comitato di dirimere le controversie che possono sorgere tra le diverse Autorità, adottando una decisione giuridicamente vincolante nelle ipotesi espressamente previste al par. 1, e nello specifico: lett. a) qualora, nell’ambito dell’attuazione del principio dell’one stop shop, vi sia un contenzioso sul progetto delineato dalla Lead Authority, in quanto l’obiezione «pertinente e motivata» sollevata da un’autorità di controllo interessata avverso il progetto di decisione dell’autorità capofila (così come descritto all’art. 60) sia da questa ritenuta non pertinente o non motivata, oppure ove l’autorità capofila non dia seguito all’obiezione; lett. b) vi sia un contrasto di opinioni tra le Autorità di controllo interessato per l’individuazione dello “stabilimento principale”, specie nell’art. 56, relativo al meccanismo dell’Autorità di controllo capofila; lett. c) un’autorità di controllo competente non abbia richiesto il parere del Comitato nei casi di cui all’articolo 64, paragrafo 1, o non si conformi (in tutto o in parte) ad esso.

L’art. 65, par. 2 delinea altresì il procedimento cui il Comitato deve conformarsi nell’adozione della “decisione vincolante”. Nello specifico si prevede che, entro un mese dal deferimento della questione da parte dell’autorità interessata (o capofila, nell’ipotesi di cui alla lett. a), il Comitato dovrà adottare una decisione vincolante, approvata dalla maggioranza qualificata di due terzi dei membri del Comitato. Il provvedimento finale del Comitato adottato deve essere motivato e trasmesso sia all’autorità che ha sollevato la questione, sia a tutte le altre autorità di controllo interessate.

A seguito della notifica della decisione vincolante da parte del Comitato, l’Autorità controllo – o, se del caso, l’autorità capofila – è tenuta ad adottare, senza ingiustificato ritardo e al più tardi entro un mese dalla ricezione della notifica, la sua decisione definitiva, la quale deve

basarsi sulla decisione vincolante del Comitato, conformandosi ad essa e facendone espresso riferimento.

Dall'analisi di queste ipotesi si comprende come la risoluzione delle controversie da parte del Comitato, configuri uno strumento di ultima istanza che interviene dopo una serie di fisiologiche fasi progressive (l'applicazione dei meccanismi di cooperazione ai sensi degli artt. 60-62, e l'emissione di un parere ai sensi dell'art. 64), delineate dal Regolamento, che avrebbero già dovuto condurre al raggiungimento del consenso tra i soggetti coinvolti .

Il par. 4, art. 65 specifica poi che, nelle more dell'adozione della decisione da parte del Comitato, le Autorità devono astenersi dall'adottare misure sulla questione sottoposta al meccanismo di coerenza, per evitare di vanificare lo sforzo del Comitato con misure potenzialmente problematiche.

Ai sensi dell'art. 66 GDPR, è tuttavia possibile sopperire ai meccanismi sinora descritti (di cooperazione e di coerenza) a causa del sopravvenire di un'urgenza. In circostanze eccezionali, infatti, ove sussista un rischio grave e immediato per i diritti e le libertà degli interessati, le Autorità che ritengano urgente provvedere, possono: a) derogare al meccanismo di cooperazione, adottando immediatamente misure provvisorie, le quali produrranno effetti giuridici unicamente nel proprio territorio per un periodo non superiore ai 3 mesi, di cui ne danno comunicazione alle altre Autorità interessate, all'EDPB e alla Commissione; b) richiedere al Comitato pareri e decisioni vincolanti d'urgenza, ossia i medesimi atti disciplinati rispettivamente dagli artt. 64 e 65, che tuttavia fruiscono di una particolare procedura accelerata, in quanto richiedono l'approvazione a maggioranza semplice (degli aventi diritto al voto all'interno del Comitato) e nel più breve termine di due settimane.

## **Conclusione e attualità**

Come visto sinora, la c.d. rete europea dei Garanti per la protezione dei dati è stata costruita dal legislatore europeo al fine di evitare l'instaurarsi delle geometrie variabili nei diversi Stati membri (ossia evitare che il Regolamento sia interpretato ed applicato diversamente negli Stati membri) e di assicurare l'omogeneità dei trattamenti in una prospettiva di armonizzazione del diritto alla protezione dei dati personali e di funzionamento del mercato interno su tutto il territorio dell'Unione. Il modello del legislatore europeo, disegnato in considerazione del bene oggetto di tutela, non è esente da criticità in quanto all'istituzionalizzazione di forme di collaborazione tra le autorità di controllo, tramite i meccanismi di cooperazione e coerenza, non ha fatto seguito l'armonizzazione di procedure amministrative e di interpretazioni dei concetti nell'ambito di tali meccanismi con la conseguenza che le singole autorità nazionali rappresentano ancora i principali attori responsabili dell'esecuzione della disciplina europea sulla tutela dei dati personali, dando luogo ad un federalismo esecutivo , da cui derivano effetti pregiudizievoli all'efficace tutela del diritto . L'applicazione coerente del GDPR dipende infatti dall'efficace funzionamento del

sistema preposto alla sua applicazione transfrontaliera . Tuttavia, benché l'EDPB abbia adottato delle Linee guida sull'applicazione dell'articolo 60 al fine di «fornire orientamenti sull'applicazione concreta delle disposizioni» relative alla procedura di cooperazione e sull'applicazione dell'art. 65, paragrafo 1, lettera a), volte a chiarire «il quadro giuridico applicabile e le fasi principali della procedura» , le differenze procedurali, che continuano a caratterizzare l'operato delle autorità di protezione dei dati, ostacolano il regolare ed effettivo funzionamento dei meccanismi di cooperazione e di composizione delle controversie del GDPR nei casi transfrontalieri, rimettendo il concreto utilizzo delle forme di coordinamento alla determinazione delle diverse autorità nazionali coinvolte .

Per far fronte a tali difficoltà e sulla base delle criticità rilevate dal Comitato , nel 2023 la Commissione ha elaborato una proposta di Regolamento contenente norme aggiuntive all'applicazione del GDPR , prodromiche a consentire la convergenza degli approcci procedurali adottati dalle autorità di protezione dei dati e rendere più efficiente e armonizzata la gestione dei casi transfrontalieri in tutta l'UE. La proposta di Regolamento è dunque tesa a definire più chiaramente le regole in diversi settori, tra cui i reclami transfrontalieri, il coinvolgimento del reclamante, i diritti della difesa delle parti oggetto dell'indagine (titolari del trattamento e responsabili del trattamento) e, soprattutto, la cooperazione tra le autorità di protezione dei dati . Pertanto, in merito ai reclami, la proposta fornisce un modulo che specifica le informazioni richieste per tutti i reclami ai sensi dell'articolo 77 GDPR relativi al trattamento transfrontaliero, al fine di evitare che le autorità di protezione interpretino in modo diverso le prescrizioni relative alla forma di un reclamo, al coinvolgimento dei reclamanti nella procedura e al rigetto degli stessi, e dunque che il trattamento dei reclami vari a seconda del luogo in cui il reclamo viene proposto o di quale autorità di protezione dei dati sia l'autorità capofila per un determinato caso.

Per quanto attiene la procedura di cooperazione di cui all'articolo 60 GDPR, invece, la proposta di Regolamento si occupa di rafforzare la sinergia tra le Autorità nei casi transfrontalieri, avendo dimostrato la recente esperienza nell'applicazione del GDPR in tali ipotesi l'insufficiente coordinamento tra autorità dei dati prima della presentazione di un progetto di decisione dell'autorità capofila, dal momento che la procedura di cui all'art. 60 è descritta a grandi linee, prevedendo unicamente che le autorità di protezione dei dati sono tenute a scambiarsi le “informazioni utili” nell'adoperarsi per raggiungere un consenso, salvo successivamente poter sollevare “obiezioni pertinenti e motivate” al progetto di decisione dell'autorità capofila. Tali obiezioni, data anche la differenza di forma e struttura con cui vengono proposte e l'assenza di un coordinamento ex ante tra le Autorità, aumentano però la probabilità di ricorso al meccanismo di composizione delle controversie di cui all'art. 65 GDPR, che, sebbene rappresenti un elemento essenziale per garantire un'interpretazione coerente della normativa privacy, dovrebbe essere riservato a casi eccezionali. La proposta dunque delinea strumenti per raggiungere più agevolmente un consenso condiviso intorno alla misura da adottare nei casi transfrontalieri, assegnando maggior concretezza all'obbligo di cooperazione e condivisione di informazioni utili da parte di autorità di controllo interessate, specificando i documenti compresi nelle informazioni che le autorità sono tenute

a condividere nel corso dell'operazione transfrontaliera e consentendogli di comunicare le loro opinioni già nelle prime fasi della procedura di indagine, in modo da incidere in modo significativo sull'indagine in corso, utilizzando anche tutti gli strumenti previsti dal GDPR, quali le indagini congiunte e l'assistenza reciproca. Verrebbe così ridotta la probabilità di divergenze nelle fasi successive della procedura che richiederebbero il ricorso al meccanismo di composizione delle controversie.

Il Regolamento appena esaminato – ancora presente allo stato di mera proposta – mette in luce gli aspetti critici della rete europea dei garanti, tentando, al contempo, di porre rimedio alla tendenza delle Autorità nazionali di ricorrere ad interpretazioni giuridiche per “eludere” il principio di cooperazione con il fine ultimo di attrarre innanzi a sé la competenza per poter garantire più velocemente la tutela dei “propri” interessati. L'effettivo rafforzamento della rete europea garantirebbe infatti maggior tutela dei diritti e il conseguente recupero della fiducia nelle istituzioni europee circa la capacità di regolare efficacemente l'evoluzione digitale.

Come ricorda il Regolamento al Considerando n.7, per far fronte alle sfide che la rapida evoluzione tecnologica e la globalizzazione presentano per la protezione dei dati personali, è più che mai indispensabile «un quadro più solido e coerente» a tutela della privacy nell'Unione, «affiancato da efficaci misure di attuazione», e che «la certezza giuridica e operativa sia rafforzata tanto per le persone fisiche quanto per gli operatori economici e le autorità pubbliche». A tal fine, è necessario che proceda il percorso di integrazione intrapreso mediante l'implementazione dei meccanismi amministrativi.

## Bibliografia

BALBONI P.– PELINO E. – SCUDIERO L., Rethinking the one-stop-shop mechanism: Legal certainty and legitimate expectation, in Computer Law & Security Review, vol. 30, n. 4/2014.

BISTOLFI C., I soggetti di controllo e verifica, in C. BISTOLFI – L. BOLOGNINI – E. PELINO (a cura di), Il Regolamento privacy europeo, Milano, Giuffrè, 2016.

BUTTARELLI G., The EU GDPR as a clarion call for a new global digital gold standard, in International Data Privacy Law, vol. 6, n. 2/2016.

CARDARELLI F., Artt. 60-62, in R. D'ORAZIO – G. FINOCCHIARO – O. POLLICINO – G. RESTA (a cura di), Codice della privacy e Data Protection, cit.

CUFFARO V., Il diritto europeo sul trattamento dei dati personali e la sua applicazione in Italia: elementi per un bilancio ventennale, in V. CUFFARO – R. D'ORAZIO – V. RICCIUTO (a cura di), I dati personali nel diritto europeo, Giappichelli, 2019.

DEL FEDERICO C.– POPOLI A. R., Disposizioni generali, in G. FINOCCHIARO (opera diretta da), Il nuovo Regolamento europeo sulla privacy e sulla protezione dei dati personali, cit.

DELLA MORTE G., Art. 70, in R. D'ORAZIO – G. FINOCCHIARO – O. POLLICINO – G. RESTA (a cura di), Codice della privacy e Data Protection, Milano, Giuffrè, 2021.

ESPOSITO M. S., Il trattamento transfrontaliero e la cooperazione tra Autorità Garanti, in G. FINOCCHIARO (opera diretta da), Il nuovo Regolamento europeo sulla privacy e sulla protezione dei dati personali, Bologna, Zanichelli, 2017.

FINOCCHIARO G., Introduzione al Regolamento Europeo sulla protezione dei dati, in Nuove leggi civ., 2017.

GONZÁLEZ FUSTER G., The Emergence of Personal Data Protection as a Fundamental Right of the EU, Bruxelles, Springer, 2014.

GUARDIGLI E., Il Garante per la protezione dei dati e la cooperazione fra Autorità Garanti, in G. FINOCCHIARO (opera diretta da), Il nuovo Regolamento europeo sulla privacy e sulla protezione dei dati personali, cit.

GUARDIGLI E., Il Garante per la protezione dei dati e la cooperazione fra Autorità Garanti, cit., 511; in proposito v. anche V. RIZZO, Artt. 55-56, in G.M. RICCIO – G. SCORZA – E. BELISARIO (a cura di), GDPR e normativa privacy. Commentario, Milano-Vicenza, Wolters Kluwer, 2018.

IPPOLITI MARTINI V. C., Comitato europeo per la protezione dei dati, in G. FINOCCHIARO (opera diretta da), Il nuovo Regolamento europeo sulla privacy e sulla protezione dei dati personali, cit.,

LUBIAN F., Art. 65, in R. D'ORAZIO – G. FINOCCHIARO – O. POLLICINO – G. RESTA (a cura di), Codice della privacy e Data Protection, cit.

LUCCHINI GUASTALLA E., Il nuovo regolamento europeo sul trattamento dei dati personali: i principi ispiratori, in Contratto e impresa, n. 1/2018.

MACCHIA M.– FIGLIOLIA C., Autorità per la privacy e Comitato europeo nel quadro del General Data Protection Regulation, in Giornale di diritto amministrativo, n. 4/2018.

PARODO F., La tutela del diritto alla protezione dei dati personali: l'effettività dei rimedi e il ruolo nomofilattico del Comitato europeo per la protezione dei dati personali, in Federalismi, n. 25/2021.

PASSAGLIA P., Privacy e nuove tecnologie, un rapporto difficile. Il caso emblematico dei social media, tra regole generali e ricerca di una specificità, in Consulta Online, III, 2016.

PIZZETTI F., Privacy e diritto europeo alla protezione dei dati personali, I e II, Torino, 2016.

PIZZETTI F., Privacy e il diritto europeo alla protezione dei dati personali, 1, Dalla Direttiva 95/46 al nuovo Regolamento europeo, Giappichelli, Torino, 2016.

PIZZETTI F., Privacy e il diritto europeo alla protezione dei dati personali, 2, Il regolamento europeo 2016/679, Giappichelli, Torino, 2016.

RATTI M., Il regime sanzionatorio previsto dal Regolamento per l'illecito dei trattamenti dei dati personali, in G. FINOCCHIARO (opera diretta da), Il nuovo Regolamento europeo sulla privacy e sulla protezione dei dati personali, cit.

SARTORETTI C., Il regolamento europeo sulla privacy: confini, sovranità e sicurezza al tempo del web, in Federalismi, n. 13/2019.

## **La sanità digitale e le sfide della coesione territoriale e dell'interoperabilità dei sistemi digitalizzati – di Beatrice Perego**

Dottessa di ricerca presso l'Università degli Studi Roma Tre, dipartimento di Scienze Politiche

La necessità di conciliare la crescente domanda di salute con la disponibilità di risorse limitate richiede un ripensamento dell'assistenza sanitaria, orientato all'impiego delle tecnologie digitali come strumento per affrontare le sfide dell'erogazione dei servizi e rafforzare il diritto alla salute<sup>208</sup>. La digitalizzazione del settore sanitario, lungi dal rappresentare una mera modernizzazione, consente un accesso più ampio, rapido ed efficiente alle prestazioni, sostenendo la transizione verso un modello proattivo, preventivo e personalizzato di cura<sup>209</sup>.

Durante la pandemia di Covid-19, la necessità di mantenere l'erogazione dei servizi ha spinto verso un'accelerazione senza precedenti dei processi digitali, evidenziando le potenzialità della telemedicina, delle prescrizioni elettroniche e di strumenti digitali di monitoraggio<sup>210</sup>. In tale contesto, l'Unione europea ha promosso il programma Next Generation EU (NGEU), finalizzato alla resilienza e alla digitalizzazione degli Stati membri, includendo la sanità tra i settori prioritari, da attuare attraverso i rispettivi *National Recovery and Resilience Plan 2021-23*<sup>211</sup>.

A livello interno, agli obiettivi previsti dal PNRR<sup>212</sup> si sono affiancati alle previsioni contenute nel Piano Triennale per l'informatica nella Pubblica Amministrazione 2020-2022, concernenti (tra gli altri) l'implementazione del Fascicolo sanitario elettronico ed i progetti di telemedicina<sup>213</sup>.

Questi strumenti, quindi, concorrono alla creazione, implementazione e tutela della c.d. “salute digitale”, la quale si riferisce all’«accesso sicuro dei cittadini ai dati sanitari e condivisione transfrontaliera degli stessi; dati migliori per far progredire la ricerca, la prevenzione delle malattie e la sanità e l'assistenza personalizzate; strumenti digitali per dare maggiori poteri e autonomia ai cittadini e per un'assistenza incentrata sulle persone», per il soddisfacimento dei quali è richiesto «lo sviluppo di standard a livello di UE per la qualità,

---

<sup>208</sup> F. SACCO, *Art. 32 della costituzione italiana: diritto alla salute - come sono state realizzate e continuano a realizzarsi le riforme nella sanità*, in Centro Studi diritto sanitario, [www.dirittosanitario.net](http://www.dirittosanitario.net) (03/04/2025).

<sup>209</sup> Cfr. VII Rapporto RBM - Censis sulla Sanità Pubblica, Privata e Intermediata. Il futuro del Sistema Sanitario in Italia tra universalismo, nuovi bisogni di cura e sostenibilità, 2017; P. DUNLEAVY, H. MARGETTS, S. BASTOW, J. TINKLER, *New public management is dead - Long live digital-era governance*, in *J Public Adm. Res Theory*, 2006, 467-494; J. BARLOW, D. SINGH, S. BAYER, R. CURRY, *A systematic review of the benefits of home telecare for frail elderly people and those with long-term conditions*, in *J. Telemed. Telecare*, 2007, 172-179.

<sup>210</sup> C. BERARDI, M. ANTONINI, M.G. GENIE, G. COTUGNO, A. LANTERI, A. MELIA, et al., *The COVID-19 pandemic in Italy: policy and technology impact on health and non-health outcomes*, in *Health Policy Technol*, 2020, 454-487.

<sup>211</sup> S. CIVITARESE MATTEUCCI, *La riforma della pubblica amministrazione nel quadro del “Recovery Fund”*, in Forum AIPDA Next Generation EU, in [www.aipda.it](http://www.aipda.it) (10/06/2023).

<sup>212</sup> Presidenza del Consiglio dei ministri, PNRR, gli obiettivi e la struttura, novembre 2021.

<sup>213</sup> AGID, Piano Triennale per l'informatica nella Pubblica Amministrazione 2020-2022, luglio 2020.

*l'affidabilità e la cibersicurezza dei dati; una standardizzazione a livello di UE delle cartelle cliniche elettroniche; e una migliore interoperabilità attraverso formati di scambio aperti»<sup>214</sup>.*

La digitalizzazione, quindi, se progettata e implementata secondo principi di equità, sicurezza e trasparenza, può contribuire alla promozione del benessere collettivo. Tuttavia, essa solleva rilevanti criticità, tra cui il rischio di depersonalizzazione della relazione di cura, la frammentazione normativa, le disuguaglianze di accesso e le vulnerabilità in materia di cybersicurezza. Inoltre, l'armonizzazione dei quadri regolatori e la definizione di standard condivisi costituiscono prerequisiti essenziali per uno sviluppo sostenibile dell'ecosistema digitale sanitario, che deve anche rispondere alle esigenze di protezione dei dati e inclusione territoriale<sup>215</sup>.

### **La Telemedicina come leva di prossimità e deospedalizzazione**

Le inefficienze associate alle valutazioni tradizionali dello stato di salute evidenziano l'importanza di esplorare approcci alternativi per far fronte, in primo luogo, alle difficoltà legate all'accesso<sup>216</sup> e ai costi delle prestazioni<sup>217</sup>; ai ritardi nei risultati analitici<sup>218</sup>; e infine, con riferimento ai tempi di attesa delle prestazioni ambulatoriali e alle brevi interazioni medico-paziente che possono non fornire una comprensione completa dello stato di salute<sup>219</sup>.

In risposta a queste criticità, l'impiego delle tecnologie digitali per supportare la pratica clinica sembrerebbe apportare significative migliorie in termini di esiti diagnostici, di efficacia e riduzione del carico di lavoro degli operatori sanitari, nonché con riferimento all'inclusione di professionisti e pazienti delle aree remote<sup>220</sup>.

Sebbene la telemedicina fosse una realtà già consolidata in alcuni Paesi anche in precedenza<sup>221</sup>, l'assistenza clinica remota è aumentata durante la pandemia di COVID-19<sup>222</sup>, in occasione della quale le Organizzazioni internazionali, come l'OMS e la Commissione Europea, hanno avviato numerose iniziative per promuoverne lo sviluppo<sup>223</sup>.

---

<sup>214</sup> Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato Economico e Sociale europeo e al Comitato delle Regioni relativa alla trasformazione digitale e dell'assistenza nel mercato unico digitale, alla responsabilizzazione dei cittadini e alla creazione di una società più sana, 25 aprile 2018.

<sup>215</sup> DI COSTANZO, *L'impiego delle nuove tecnologie nel settore della salute: problematiche e prospettive di diritto costituzionale*, in *Consulta Online*, 2023, 214 e ss.

<sup>216</sup> AGENAS, *Valutazione della performance delle aziende ospedaliere pubbliche*, vol. 49, in *Monitor elementi di analisi e osservazioni del sistema salute*, 2024, 35 e ss.

<sup>217</sup> Commissione europea, *State of Health in the EU - Italia: Profilo della sanità*, 2023, 18.

<sup>218</sup> *Ibidem*, 13.

<sup>219</sup> AGENAS, *Monitoraggio ex-ante dei tempi di attesa delle prestazioni ambulatoriali – anno 2023 – attività di sperimentazione*, 2023,

[https://www.agenas.gov.it/images/2023/10\\_nov/Comunicato\\_Sperimentazione\\_Monitoraggio\\_Tempi\\_di\\_Atesa\\_Erogazione\\_prestazioni\\_I\\_semestre\\_2023.pdf](https://www.agenas.gov.it/images/2023/10_nov/Comunicato_Sperimentazione_Monitoraggio_Tempi_di_Atesa_Erogazione_prestazioni_I_semestre_2023.pdf) (05/04/2024); AGENAS, *Valutazione della performance delle aziende ospedaliere pubbliche*, in *Monitor elementi di analisi e osservazioni del sistema salute*, 2024, 35 e ss.

<sup>220</sup> The Lancet Gastroenterology Hepatology, *The potential of telemedicine in digestive diseases*, in *Lancet Gastroenterol Hepatol*, 2019; D.L. ARMAIGNAC, ET AL., *Impact of telemedicine on mortality, length of stay, and cost among patients in progressive care units: experience from a large healthcare system*, in *Crit Care Med*, 2018.

<sup>221</sup> M. SURAN, *Increased use of medicare telehealth during the pandemic*, in *JAMA*, 2022.

<sup>222</sup> *Ibidem*.

<sup>223</sup> European Commission, *Europe fit for the digital age: Empowering people with a new generation of technologies*, [https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age\\_en](https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age_en) (07/04/2025)

Ne è segnale anche la pubblicazione in Italia delle Linee Guida per la telemedicina<sup>224</sup>, che si inseriscono tra gli interventi inclusi nel PNRR<sup>225</sup> e definiscono un modello di riferimento per l'attuazione dei diversi servizi di telemedicina nel *setting* domiciliare, attraverso l'individuazione di processi innovativi di presa in carico del paziente a domicilio e la valorizzazione della collaborazione multidisciplinare tra i diversi professionisti<sup>226</sup>.

Inoltre, l'Agenzia Nazionale per i Servizi Sanitari Regionali (AGENAS), in qualità di soggetto attuatore, ha dato seguito alla realizzazione della Piattaforma Nazionale di telemedicina (PNT), che nasce con l'obiettivo di potenziare l'assistenza territoriale, ridurre le disparità nell'accesso ai servizi sanitari, facilitare la gestione dei pazienti acuti e cronici e integrare i servizi sanitari regionali con le piattaforme nazionali<sup>227</sup>.

La PNT si articola nell'Infrastruttura Nazionale di Telemedicina (INT) gestita da AGENAS, che garantisce l'interoperabilità e il monitoraggio a livello nazionale, e le Infrastrutture Regionali di Telemedicina (IRT) amministrate dalle Regioni e dalle Province autonome, responsabili dell'erogazione diretta dei servizi di telemedicina<sup>228</sup>.

Le linee guida operative stabiliscono poi i criteri per l'eleggibilità dei pazienti ai servizi di telemedicina, valutando non solo le condizioni cliniche, ma anche gli aspetti tecnologici, culturali e di supporto familiare. Particolare attenzione è rivolta alla *digital literacy* e alla disponibilità di dispositivi adeguati all'accesso alle prestazioni<sup>229</sup>. In questo senso, la telemedicina non sostituisce le prestazioni in presenza, ma le integra, estendendone il campo di azione e l'ambito temporale.

## Il Fascicolo Sanitario Elettronico e la gestione dei dati clinici

Il presupposto per ottenere un modello efficiente e di miglior accesso alle cure è rappresentato dalla creazione di un ambiente digitale integrato che permetta un'analisi costante dei dati dei pazienti<sup>230</sup>.

---

<sup>224</sup> Linee Guida organizzative contenenti il modello digitale per l'attuazione dell'assistenza domiciliare (Milestone EU M6C1-4), [https://www.salute.gov.it/imgs/C\\_17\\_pagineAree\\_5874\\_0\\_file.pdf](https://www.salute.gov.it/imgs/C_17_pagineAree_5874_0_file.pdf) (07/04/2025); Ministero della Salute, *Decreto 29 aprile 2022, Approvazione delle linee guida organizzative contenenti il «Modello digitale per l'attuazione dell'assistenza domiciliare», ai fini del raggiungimento della Milestone EU M6C1-4, di cui all'Annex alla decisione di esecuzione del Consiglio ECOFIN del 13 luglio 2021, recante l'approvazione della valutazione del Piano per la ripresa e resilienza dell'Italia*, G.U. Serie Generale, n. 120 del 24 maggio 2022.

<sup>225</sup> Il testo del PNRR è consultabile in [www.governo.it](http://www.governo.it) (10/04/2025).

<sup>226</sup> Gli elementi fondamentali della riorganizzazione dell'assistenza territoriale sono principalmente: la Casa della Comunità (CdC), il Medico di Medicina Generale – Pediatra di Libera scelta (MMG/PLS), la Centrale Operativa Territoriale (COT), l'Infermiere di Famiglia o di Comunità (IFoC), l'Unità di Continuità Assistenziale (UCA) e la rete di Cure palliative in cui è inserito, quale nodo della rete, anche il *setting* domiciliare. Linee Guida organizzative contenenti il modello digitale per l'attuazione dell'assistenza domiciliare, *op.cit.*, 3 e ss.

<sup>227</sup> Ministero della Salute, Agenas, *Piattaforma Nazionale di telemedicina*, <https://pnt.agenas.it/home> (10/01/2025). Si veda Garante per la protezione dei dati personali, *Parere su uno schema di decreto recante la disciplina dei trattamenti di dati personali nell'ambito della Piattaforma nazionale di telemedicina*, 16 gennaio 2025 [Doc. web n. 10105743].

<sup>228</sup> *Ibidem*.

<sup>229</sup> Linee Guida organizzative contenenti il modello digitale per l'attuazione dell'assistenza domiciliare (Milestone EU M6C1-4), [https://www.salute.gov.it/imgs/C\\_17\\_pagineAree\\_5874\\_0\\_file.pdf](https://www.salute.gov.it/imgs/C_17_pagineAree_5874_0_file.pdf) (07/04/2025).

<sup>230</sup> S.J. MOONEY, V. PEJAVER, *Big Data in Public Health: Terminology, Machine Learning, and Privacy*, in *Annu Rev Public Health*, 2018, 2.

La gestione dei dati clinici su larga scala, infatti, ha assunto un ruolo sempre più rilevante nella promozione della salute della popolazione<sup>231</sup>. In tale contesto, il processo di raccolta e accesso alle informazioni necessarie a ricostruire la storia clinica dei pazienti<sup>232</sup> è stato velocizzato e semplificato dall'impiego da strumenti tecnologici<sup>233</sup> tra cui il Fascicolo Sanitario Elettronico, che si è rivelato uno dei pilastri della digitalizzazione del settore sanitario.

Il FSE, rappresentando un archivio digitale completo delle informazioni sanitarie di un paziente, permette ai professionisti sanitari fornire cure efficienti e personalizzate, riducendo i tempi di attesa e consentendo una migliore presa di decisione clinica<sup>234</sup>.

In Italia, ai sensi dell'art. 12, co. 2, d.l. n. 179/2012, l'istituzione del FSE è prevista «*nel rispetto della normativa vigente in materia di protezione dei dati personali, ai fini di: a) prevenzione, diagnosi, cura e riabilitazione; b) studio e ricerca scientifica in campo medico, biomedico ed epidemiologico; c) programmazione sanitaria, verifica delle qualità delle cure e valutazione dell'assistenza sanitaria»* e la consultazione dei dati e documenti in esso contenuti, per le finalità di cui alla lett. a), «può essere realizzata soltanto con il consenso dell'assistito e sempre nel rispetto del segreto professionale, salvo i casi di emergenza sanitaria secondo modalità individuate a riguardo», mentre le finalità di cui alle lett. b) e c) sono perseguiti da Stato, Regioni e Province autonome «*in conformità ai principi di proporzionalità, necessità e indispensabilità nel trattamento dei dati personali»*<sup>235</sup>.

Con il decreto ministeriale del 20 maggio 2022<sup>236</sup> sono state poi pubblicate le Linee Guida per l'Attuazione dello stesso<sup>237</sup>, le quali veicolano la progettazione e lo sviluppo di un sistema uniforme sull'intero territorio nazionale<sup>238</sup>.

Ora, sebbene il FSE si sia rivelato una risorsa preziosa ai fini dell'ottimizzazione dell'accesso alle informazioni sanitarie e della coordinazione delle cure, l'Autorità Garante per la protezione dei dati personali ha messo in luce una serie di criticità in occasione del parere negativo del 22 agosto 2022, espresso sulla base dello schema di decreto del Ministero della salute volto a dare una prima attuazione alle modifiche dell'art. 12 del decreto-legge 18 ottobre 2012, n. 179 sul FSE<sup>239</sup>.

Nello specifico, venivano rappresentate numerose carenze strutturali e sostanziali, riguardanti la mancanza di garanzie uniformi a livello nazionale e degli elementi richiesti dallo stesso art. 12, rilevando l'esigenza di riformulare lo schema di decreto<sup>240</sup> e sottolineando

---

<sup>231</sup> European Commission, *Directorate-General for Health and Consumers, Unit D3 eHealth and Health Technology Assessment. The use of big data in public health policy and research, background information document*, 2014.

<sup>232</sup> M. FARINA, *Il cloud computing in ambito sanitario tra security e privacy*, Milano, 2019, 43.

<sup>233</sup> *Ibidem*.

<sup>234</sup> E. SORRENTINO, A.F. SPAGNUOLO, *La sanità digitale in emergenza Covid-19. Uno sguardo al fascicolo sanitario elettronico*, in *federalismi.it*, 2020.

<sup>235</sup> L. CALIFANO, *Fascicolo sanitario elettronico (Fse) e dossier sanitario. Il contributo del Garante privacy al bilanciamento tra diritto alla salute e diritto alla protezione dei dati personali*, in *San. pubb. e priv.*, 2015, 7 ss.

<sup>236</sup> Ministero della Salute - Decreto ministeriale 20 maggio 2022 recante *Adozione delle Linee guida per l'attuazione del Fascicolo sanitario elettronico*.

<sup>237</sup> In particolare, ci si riferisce all'investimento 1.3.1, Missione 6, Componente 2. Piano Nazionale di Ripresa e Resilienza, p. 231 e ss., [www.governo.it](http://www.governo.it) (10/04/2024).

<sup>238</sup> Cfr Ministero della Salute - Decreto ministeriale 18 maggio 2022. *Integrazione dei dati essenziali che compongono i documenti del Fascicolo sanitario elettronico*

<sup>239</sup> Garante per la protezione dei dati personali, Parere al Ministero della Salute sullo schema di decreto, da adottare assieme al Ministro delegato per l'innovazione tecnologica e la transizione digitale, di concerto con il Ministro dell'economia e delle finanze, sul Fascicolo Sanitario Elettronico (FSE), n. 294, 22 agosto 2022.

<sup>240</sup> Garante privacy, parere 22 agosto 2022, *op. cit.*, 34 e ss.

l'importanza di garantire privacy e sicurezza dei dati<sup>241</sup>, la necessità di sistemi di raccolta standardizzati e protocolli di verifica<sup>242</sup>, la scarsa interoperabilità tra i diversi sistemi informatici<sup>243</sup> ed infine, l'importanza del coinvolgimento attivo dei pazienti, al fine di garantire trasparenza e rafforzare la fiducia nel sistema sanitario digitale<sup>244</sup>.

Tali principi sono stati ribaditi dal Garante con riferimento alle prescrizioni contenute nel decreto del 7 settembre 2023 relativo al Fascicolo Sanitario Elettronico 2.0.<sup>245</sup>, che ha messo in luce le numerose violazioni riscontrate nell'attuazione della nuova disciplina<sup>246</sup>, evidenziando come «*alcuni diritti [...] e misure [...] non sono garantite in modo uniforme in tutto il Paese. Oppure sono esercitabili ed esigibili solo dagli assistiti di talune Regioni e Province autonome, con un potenziale e significativo effetto discriminatorio sugli assistiti*»<sup>247</sup>.

In definitiva, pur dovendo far fronte agli ostacoli evidenziati, il FSE rappresenta un elemento chiave nella trasformazione digitale della sanità moderna, contribuendo a migliorare la qualità delle cure, ridurre gli errori medici e ottimizzare l'allocazione delle risorse sanitarie e l'accesso ai servizi offerti.

## L'Ecosistema dei Dati Sanitari Nazionale come strumento di integrazione

Nel percorso di trasformazione digitale della sanità italiana promosso dal PNRR, l'integrazione tra FSE e Telemedicina rappresenta uno snodo strategico fondamentale per rendere i servizi sanitari più moderni, accessibili ed efficienti<sup>248</sup>.

Nello specifico, l'interoperabilità tra le due piattaforme si realizza attraverso l'Ecosistema Dati Sanitari (EDS), che funge da nodo comune per la raccolta, la validazione e lo scambio dei dati e degli eventi clinici<sup>249</sup>.

Così, ad esempio, un referto generato durante una televisita viene automaticamente validato, convertito nello standard FHIR<sup>250</sup>, memorizzato nel repository dell'EDS e reso disponibile nel FSE per la consultazione<sup>251</sup>. Allo stesso tempo, i dati di monitoraggio continuo o

<sup>241</sup> *Ibidem*, 10 e ss.

<sup>242</sup> *Ibidem*, 8 e ss.

<sup>243</sup> *Ibidem*, 26 e ss. Sul punto si vedano Proposta di Regolamento del Parlamento Europeo e del Consiglio sullo spazio europeo dei dati sanitari, 3 maggio 2022, [20220407-EHDS for Decide \(003\) \(europa.eu\)](#) (14/04/2024); Parere congiunto 1/2019 dell'European Data Protection Board e dell'European Data Protection Supervisor sul trattamento dei dati dei pazienti e sul ruolo della Commissione europea nell'ambito dell'infrastruttura di servizi digitali per la sanità elettronica (eHDSI), [19-07-15 edpb edps joint opinion ehealth\\_en.pdf \(europa.eu\)](#) (14/04/2024), §§ 72 e 77.

<sup>244</sup> Garante privacy, parere 22 agosto 2022, *op. cit.*, 16 e ss.

<sup>245</sup> Ministero della Salute, Decreto 7 settembre 2023, *Fascicolo sanitario elettronico 2.0.*, GU n. 249 del 24 ottobre 2023.

<sup>246</sup> Garante per la protezione dei dati personali, *FSE 2.0, Garante Privacy: al via procedimenti nei confronti di 18 Regioni e 2 Province autonome. La grave situazione segnalata al Presidente del Consiglio e al Ministro della salute nei giorni scorsi*, Newsletter n. 525 del 26 giugno 2024.

<sup>247</sup> *Ibidem*.

<sup>248</sup> G. LOFARO, *Piattaforma di Telemedicina e Fascicolo Sanitario Elettronico: il accordo dei flussi informativi per i servizi sanitari digitali alla luce delle nuove linee guida*, in *Amministrativamente.com*, 2023.

<sup>249</sup> L'Ecosistema Dati Sanitari è stato istituito ai sensi del comma 15-quater dell'articolo 12, comma 15-quater del decreto-legge 18 ottobre 2012, n. 179, convertito, con modificazioni, dalla legge 17 dicembre 2012, n. 221, e successive modificazioni.

<sup>250</sup> R. SARIPALLE, C. RUNYAN, M. RUSSELL, *Using HL7 FHIR to achieve interoperability in patient health record*, in *Journal of biomedical informatics*, 2019, 94.

<sup>251</sup> G. LOFARO, *op. cit.*, 897 e ss.

gli eventi clinici vengono gestiti e condivisi tra le piattaforme, abilitando una presa in carico più tempestiva e personalizzata<sup>252</sup>.

Questa struttura consente ai cittadini di accedere in modo fluido e trasparente ai propri dati sanitari, anche prodotti a distanza, mentre i professionisti possono disporre di un quadro clinico completo e aggiornato, migliorando diagnosi, continuità assistenziale e prevenzione<sup>253</sup>.

Uno degli obiettivi centrali dell'EDS, inoltre, è quello di favorire la coesione territoriale: l'architettura federata, unita all'adozione di formati e terminologie comuni, consente di creare un ambiente interoperabile in cui ogni assistito può beneficiare della stessa qualità informativa, a prescindere dalla Regione di appartenenza<sup>254</sup>.

L'interoperabilità si declina così su più livelli: quello tecnico, attraverso l'uso dello standard HL7-FHIR; quello “semantico”, grazie all'impiego di vocabolari clinici condivisi; quello organizzativo, con workflow armonizzati tra FSE, EDS e PNT; ed infine quello giuridico, mediante un quadro normativo aggiornato e coerente con il GDPR.

Ora, sebbene l'EDS si configuri oggi come uno strumento strategico, la realizzazione del suo potenziale è ancora frenata da una serie di criticità che riguardano, tra gli altri aspetti, l'eterogeneità dei dati raccolti<sup>255</sup> e la resistenza culturale da parte di operatori sanitari<sup>256</sup>.

Un ulteriore nodo riguarda la tutela della privacy: il rispetto del GDPR, la gestione dinamica del consenso e l'adozione di misure di sicurezza efficaci rappresentano sfide centrali, soprattutto nelle realtà locali meno dotate di competenze specialistiche<sup>257</sup>. A ciò si aggiunge la complessità della governance multilivello, che coinvolge attori istituzionali differenti, la cui azione coordinata è indispensabile per evitare sovrapposizioni, ritardi e inefficienze.

In conclusione, l'EDS si configura come un'infrastruttura imprescindibile per il futuro della sanità digitale in Italia e la sua efficacia dipenderà dalla capacità del sistema di armonizzare le differenze territoriali, standardizzare i processi e promuovere una cultura dell'innovazione condivisa tra tutti i soggetti coinvolti.

### Brevi note conclusive

Strumenti come la telemedicina, il FSE e l'Ecosistema dei Dati Sanitari possono abilitare una nuova governance del sistema salute, più equa, tempestiva, personalizzata e sostenibile. Tuttavia, questa trasformazione non è automatica, richiedendo, al contrario, investimenti in infrastrutture, formazione, innovazione organizzativa e soprattutto interoperabilità.

---

<sup>252</sup> *Ibidem*.

<sup>253</sup> F. F. PAGANO, *Pubblica Amministrazione e innovazione tecnologica*, in *La Rivista “Gruppo di Pisa”*, 2021, p. 84 e ss.

<sup>254</sup> Nello specifico si vedano p. 6; Articolo 4, comma 1, lettera a) e d) (struttura architettonale dell'EDS); Articolo 14 (Accesso ai servizi per finalità di prevenzione); Articolo 19 (Monitoraggio e controllo).

<sup>255</sup> A. BENECHI, C. BOTTONI, E. CIAPANNA, A. FRIGO, A. MILAN, E. SCARINZ, *Digitalisation in Italy: evidence from a new regional index*, in *Questioni di economia e finanza (occasional papers)*, 2021, 18.

<sup>256</sup> Per una panoramica sul tema cfr. I.J. BORGES DO NASCIMENTO, ET AL., *Barriers and facilitators to utilizing digital health technologies by healthcare professionals*, in *NPJ Digit Med*, 2023. Cfr. anche World Health Organization, *Golbal Strategy on Digital Health 2020-2025*, 2021.

<sup>257</sup> Garante privacy, parere 22 agosto 2022, *op. cit.*; Garante privacy, *FSE 2.0, Garante Privacy: al via procedimenti nei confronti di 18 Regioni e 2 Province autonome*, *op. cit.*; World Health Organization, *Regulatory considerations on artificial intelligence for health*, Ginevra, 2023, 33- 39 e ss.

Invero, come si evince dai dati contenuti nel rapporto **Desi 2022**, sebbene la crisi pandemica abbia contribuito ad accelerare e implementare processi di integrazione del digitale nei servizi pubblici, il processo di digitalizzazione nel nostro paese si configura ancora al di sotto della media europea<sup>258</sup>, registrandosi, inoltre, ancora un ampio divario tra Nord e Sud<sup>259</sup>.

Ciò in quanto sono mancate in parte delle politiche di welfare in grado di dare una risposta adeguata alle fasce della popolazione maggiormente colpite dall'introduzione delle nuove tecnologie e sussiste ancora una ridotta consapevolezza rispetto all'adozione di misure di sicurezza a fini di protezione dei dati personali in ossequio alla normativa in materia<sup>260</sup>.

La scarsa efficienza dei servizi erogati on-line e delle infrastrutture e attrezzature digitali associata alla carenza di competenze digitali nel nostro Paese<sup>261</sup>, ha significative ripercussioni sul c.d. *digital divide*<sup>262</sup>, che penalizza in particolare gli anziani e coloro che vivono nelle aree più remote della penisola<sup>263</sup>.

In secondo luogo, l'assenza di un'infrastruttura coerente e interconnessa costituisce un limite strutturale al raggiungimento degli obiettivi fissati dal PNRR, aggravando le diseguaglianze tra regioni e rendendo disomogenea la qualità dei servizi offerti alla popolazione<sup>264</sup>.

Ne deriva una situazione in cui il progresso tecnologico non sempre si traduce in un effettivo miglioramento dei servizi, bensì rischia di aumentare le disparità preesistenti, generando sfiducia e resistenze culturali da parte degli operatori e dei cittadini.

Per rispondere a queste sfide è necessario un impegno politico e istituzionale forte e coordinato, orientato alla riduzione delle asimmetrie territoriali attraverso investimenti in grado di garantire la piena interoperabilità tra i sistemi regionali e assicurare a tutti i cittadini l'accesso a servizi sanitari digitali di qualità, sicuri e realmente inclusivi<sup>265</sup>.

In questo senso, la sanità digitale deve essere concepita non come un fine in sé, ma come uno strumento abilitante per l'effettività del diritto alla salute, che, in quanto diritto fondamentale, esige di essere garantito in modo uniforme e sostanziale su tutto il territorio nazionale. La sfida della coesione digitale, dunque, si intreccia indissolubilmente con quella

<sup>258</sup> Digital Economy and Society Index (DESI) - Thematic chapters, 2022, 17 e ss. Si noti che a partire dal 2023, il DESI è integrato nella relazione sullo stato del decennio digitale 2030 per monitorare i progressi compiuti rispetto al conseguimento degli obiettivi di digitalizzazione. [DESI dashboard for the Digital Decade \(2023 onwards\) - Digital Decade DESI visualisation tool](#) (20/07/2025).

<sup>259</sup> A. BENECHI, C. BOTTONI, E. CIAPANNA, A. FRIGO, A. MILAN, E. SCARINZ, *Digitalisation in Italy: evidence from a new regional index*, in *Questioni di economia e finanza (occasional papers)*, 2021, 18.

<sup>260</sup> Garante Privacy, Parere 24 febbraio 2022; R. LOMBARDI, *Sicurezza dei dati in ambito sanitario ed evoluzione tecnologica tra passato, presente e futuro*, in *Il diritto dell'economia*, 70-75.

<sup>261</sup> Commissione europea, *Indice di digitalizzazione dell'economia e della società 2022 – Italia*, 2022, 7; I.J. BORGES DO NASCIMENTO, ET AL., *Barriers and facilitators to utilizing digital health technologies by healthcare professionals*, op. cit.

<sup>262</sup> Eurostat, *Eurostat regional yearbook*, 2022; Istat, *Cittadini e ict*, 2019.

<sup>263</sup> Cfr S. LUCATELLI, *La Strategia Nazionale Aree Interne (stimolare la crescita investendo nella coesione territoriale)*, 2015; G. CARROSIO, *A place-based perspective for welfare recalibration in the Italian inner peripheries: the case of the Italian strategy for inner areas*, in *Sociol. e Polit. Soc.*, 2016, 50-64; H. YOUNG, Y. JANG, P.W. VAUGHAN, M. GARCIA, *Older Adults' internet use for health information: digital divide by race/ethnicity and socioeconomic status*, in *J. Appl. Gerontol.*, 2020, 105-110.

<sup>264</sup> Organizzazione per la Cooperazione e lo Sviluppo Economico, *The Digital Divide in Rural Healthcare*, 2020; J. BADAR, A. MOTULSKY, J.L. DENIS, *Digital health technologies and inequalities: A scoping review of potential impacts and policy recommendations*, in *Health Policy*, 2024.

<sup>265</sup> European Commission, *The future of European competitiveness*, 9 settembre 2024, 24 e ss.

della giustizia sociale, e solo un'azione consapevole e sistemica potrà consentire al sistema sanitario italiano di evolvere verso un modello più equo, resiliente e orientato al futuro.

## Bibliografia

- AGENAS, *Monitoraggio ex-ante dei tempi di attesa delle prestazioni ambulatoriali – anno 2023 – attività di sperimentazione*, 2023,  
[https://www.agenas.gov.it/images/2023/10\\_nov/Comunicato\\_Sperimentazione\\_Monitoraggio\\_Tempi\\_di\\_Atesa\\_Erogazione\\_prestazioni\\_I\\_semestre\\_2023.pdf](https://www.agenas.gov.it/images/2023/10_nov/Comunicato_Sperimentazione_Monitoraggio_Tempi_di_Atesa_Erogazione_prestazioni_I_semestre_2023.pdf) (05/04/2024).
- AGENAS, *Valutazione della performance delle aziende ospedaliere pubbliche*, vol. 49, in *Monitor elementi di analisi e osservazioni del sistema salute*, 2024.
- ARMAIGNAC D. L., ET AL., *Impact of telemedicine on mortality, length of stay, and cost among patients in progressive care units: experience from a large healthcare system*, in *Crit Care Med*, 2018.
- BARLOW J., SINGH D., BAYER S., CURRY R., *A systematic review of the benefits of home telecare for frail elderly people and those with long-term conditions*, in *J. Telemed. Telecare*, 2007.
- BENECHI A., BOTTONI C., CIAPANNA E., FRIGO A., MILAN A., SCARINZ E., *Digitalisation in Italy: evidence from a new regional index*, in *Questioni di economia e finanza (occasional papers)*, 2021.
- BENECHI A., BOTTONI C., CIAPANNA E., FRIGO A., MILAN A., SCARINZ E., *Digitalisation in Italy: evidence from a new regional index*, in *Questioni di economia e finanza (occasional papers)*, 2021.
- BERARDI C., ANTONINI M., GENIE M. G., COTUGNO G., LANTERI A., MELIA A., et al., *The COVID-19 pandemic in Italy: policy and technology impact on health and non-health outcomes*, in *Health Policy Technol*, 2020.
- BORGES DO NASCIMENTO I. J., ET AL., *Barriers and facilitators to utilizing digital health technologies by healthcare professionals*, in *NPJ Digit Med*, 2023.
- CALIFANO L., *Fascicolo sanitario elettronico (Fse) e dossier sanitario. Il contributo del Garante privacy al bilanciamento tra diritto alla salute e diritto alla protezione dei dati personali*, in *San. pubb. e priv.*, 2015.
- CARROSIO G., *A place-based perspective for welfare recalibration in the Italian inner peripheries: the case of the Italian strategy for inner areas*, in *Sociol. e Polit. Soc.*, 2016.
- CIVITARESE MATTEUCCI S., *La riforma della pubblica amministrazione nel quadro del “Recovery Fund”*, in *Forum AIPDA Next Generation EU*, in [www.aipda.it](http://www.aipda.it) (10/06/2023).
- DI COSTANZO, *L'impiego delle nuove tecnologie nel settore della salute: problematiche e prospettive di diritto costituzionale*, in *Consulta Online*, 2023.
- DUNLEAVY P., MARGETTS H., BASTOW S., TINKLER J., *New public management is dead - Long live digital-era governance*, in *J Public Adm. Res Theory*, 2006.
- FARINA M., *Il cloud computing in ambito sanitario tra security e privacy*, Milano, 2019.

LOFARO G., *Piattaforma di Telemedicina e Fascicolo Sanitario Elettronico: il raccordo dei flussi informativi per i servizi sanitari digitali alla luce delle nuove linee guida*, in Amministrativamente.com, 2023.

LOMBARDI P., *Sicurezza dei dati in ambito sanitario ed evoluzione tecnologica tra passato, presente e futuro*, in *Il diritto dell'economia*, 2021.

LUCATELLI S., *La Strategia Nazionale Aree Interne (stimolare la crescita investendo nella coesione territoriale)*, 2015.

MOONEY S. J., PEJAVER V., *Big Data in Public Health: Terminology, Machine Learning, and Privacy*, in *Annu Rev Public Health*, 2018.

PAGANO F. F., *Pubblica Amministrazione e innovazione tecnologica*, in *La Rivista “Gruppo di Pisa”*, 2021.

SACCO F., *Art. 32 della costituzione italiana: diritto alla salute - come sono state realizzate e continuano a realizzarsi le riforme nella sanità*, in Centro Studi diritto sanitario, [www.dirittosanitario.net](http://www.dirittosanitario.net) (03/04/2025).

SARIPALLE R., RUNYAN C., RUSSELL M., *Using HL7 FHIR to achieve interoperability in patient health record*, in *Journal of biomedical informatics*, 2019.

SORRENTINO E., SPAGNUOLO A. F., *La sanità digitale in emergenza Covid-19. Uno sguardo al fascicolo sanitario elettronico*, in *federalismi.it*, 2020.

SURAN M., *Increased use of medicare telehealth during the pandemic*, in *JAMA*, 2022.

YOON H., JANG Y., VAUGHN P. W., GARCIA M., *Older Adults' internet use for health information: digital divide by race/ethnicity and socioeconomic status*, in *J. Appl. Gerontol.*, 2020.





## Nextwork360 - Boosting Tech Company Growth

Nextwork360 è il network di riferimento in Italia di testate e siti B2B dedicati ai temi della Trasformazione Digitale e dell'Innovazione Imprenditoriale. Aggrega la più grande community di CIO, Innovation Manager, C-Level, imprenditori tech e ha la missione di diffondere la cultura digitale e imprenditoriale nelle imprese e pubbliche amministrazioni italiane e favorirne l'incontro con i migliori fornitori tecnologici attraverso eventi, progetti di comunicazione, lead generation e servizi inbound.

[info@nextwork360.it](mailto:info@nextwork360.it)

**Nextwork360 S.r.l.**

Via Moretto da Brescia 22, Milano

