

Sistemi per il rilevamento delle emozioni e protezione dei dati: prime osservazioni sull'interazione tra GDPR e AI Act

Emotion recognition systems and data protection: first insights on the interaction between GDPR and AI Act

LORENZO COREA (ORCID) 

Dottore di ricerca in Diritto privato e comparato, diritto del lavoro e tradizione giuridica europea
Università degli studi di Pisa

Abstract

L'articolo si propone di esaminare il perimetro della protezione dei dati personali quando vengano trattati da chatbot dotati di machine learning e possibilità di inferire le emozioni umane. L'elaborato è strutturato in una prima descrizione dell'evoluzione dei modelli di apprendimento, per poi passare alla problematizzazione delle domande di senso e delle questioni giuridiche rilevanti in materia di tutela dei diritti fondamentali, con particolare riferimento al pericolo di "datafication" della persona umana e alle prassi di profilazione algoritmica. Il contributo mette a confronto il quadro normativo del GDPR e le innovazioni introdotte dall'AI Act, ponendo particolare attenzione alle misure di trasparenza e spiegabilità in bilanciamento con gli altri diritti. L'analisi si conclude con delle riflessioni sulle prospettive evolutive e sulla necessità di tutela della dignità umana nella prospettiva del bilanciamento tra diritti.

This article aims to examine the scope of personal data protection when data is processed by chatbots equipped with machine learning and human emotion detection possibilities. The paper begins with a description of the evolution of learning models, before moving on to examine the relevant legal issues and questions of meaning in relation to the protection of fundamental human rights, with reference to the danger of "datafication" and algorithmic profiling practices. The paper compares the regulatory framework of the GDPR, and the innovations introduced by the AI Act, paying particular attention to transparency and explainability measures in balance with other rights. The analysis concludes with reflections on future developments and on the need of protection of human dignity with a view to balancing right.



Keywords: Chatbot emozionali; Intelligenza artificiale; Dati biometrici; GDPR; AI Act; Profilazione algoritmica; Diritti fondamentali.

Sommario: [1. L'evoluzione delle scienze cognitive e l'uso dei chatbot.](#) – [2. Quando l'etica si intreccia con il diritto. Domande di senso e questioni giuridicamente rilevanti.](#) – [3. AI Act & GDPR: una coppia che "funziona".](#) – [4. L'integrazione "in concreto" della normativa vigente e le misure tecniche di protezione](#) – [5. Conclusioni e prospettive evolutive.](#)

1. L'evoluzione delle scienze cognitive e l'uso dei chatbot.

Le macchine di apprendimento hanno subito una significativa evoluzione negli ultimi anni, passando da semplici sistemi di risposta automatica basati su script predefiniti a sofisticati modelli basati sull'intelligenza artificiale generativa, in grado di interagire in modo flessibile e naturale con l'utente.

Preponderante, almeno inizialmente, era l'approccio proposto da un illustre pioniere delle scienze informatiche come Alan Turing, che ideò il test di Turing partendo dalla domanda "Can a machine think?"¹. La sua risposta aveva orientato anni di dibattito attorno alla funzione dell'intelligenza artificiale, arrivando a postulare la teoria dell'"intelligenza artificiale forte", per cui un computer potrebbe ottenere gli stessi risultati di una mente umana, cioè la possibilità di pensare, di avere stati cognitivi, di capire discorsi e domande al fine di rispondere. Secondo tale concezione l'impostazione di un programma basato su simboli e regole di calcolo abilita la macchina ad elaborare gli input ricevuti attraverso un procedimento di manipolazione di simboli per emettere, conseguentemente, un output.

I chatbot di ultima generazione, basati su modelli linguistici di grandi dimensioni (Large Language Model, LLM), hanno inaugurato una nuova fase nell'interazione uomo-macchina e superato il test di Turing, in quanto specializzati e addestrati su grandi masse di dati allo scopo di eccellere in un determinato settore, quello della generazione linguistica naturale (NLP).

Significa che i modelli LLM sono intelligenti al pari dell'uomo? Per rispondere a tale interrogativo ci vengono in aiuto le confutazioni proposte da Searle con la teoria della "stanza cinese"². Il filosofo statunitense ha affermato che le

¹ A. TURING, *Computing Machinery and Intelligence*, in *Mind*, 1950, LIX, 236, 433-460. Il test di Turing, proposto nel 1950 dal matematico e informatico britannico Alan Turing, è un esperimento progettato per valutare la capacità di una macchina di mostrare un comportamento intelligente indistinguibile da quello di un essere umano. Nella formulazione classica, il test coinvolge tre partecipanti: un interrogatore umano, un rispondente umano e una macchina. L'interrogatore comunica con entrambi i partecipanti a lui non palesati tramite messaggi scritti, con l'obiettivo di determinare quale dei due sia la macchina. Se l'interrogatore non è in grado di distinguere in modo chiaro la macchina dall'essere umano, si può concludere che la macchina abbia dimostrato un'intelligenza simile a quella umana.

² J. R. SEARLE, *Minds, brains, and programs*, in *Behavioral and Brain Sciences*, 1980, III, 417-424. Searle ci chiede di immaginare una persona che non capisce il cinese seduta in una stanza chiusa. La persona riceve dei foglietti con caratteri cinesi (domande) attraverso una fessura e all'interno della stanza ha un libro di regole, scritto nella sua lingua, che le spiega esattamente come manipolare questi simboli e produrre le risposte corrette in cinese. L'individuo segue perfettamente le regole e restituisce i foglietti attraverso la fessura, producendo risposte indistinguibili da quelle di una madrelingua cinese.

Dall'esterno, sembra che la persona nella stanza comprenda e capisca il cinese. Ma in realtà, sostiene l'A., la persona sta solo manipolando i simboli secondo regole di forma sequenziali, senza alcuna comprensione "reale" del loro significato.

macchine non pensano, non sono intenzionali, ma possono al massimo emulare alcune funzioni biologiche del cervello. Egli sostiene che l'intelligenza artificiale non possa essere equivalente a quella umana perché non è sufficiente elaborare programmi di manipolazione di simboli secondo regole sintattiche per generare un'attività mentale. Il fatto è che la mente umana comprende, elabora e si esprime attraverso un linguaggio le cui parole da un lato sono investite di significato e dall'altro determinano il modo in cui una risposta verrà data, emergendo così l'esperienza soggettiva e unica dell'essere umano.

Le applicazioni odierne – da ChatGPT a Copilot, da Claude a Gemini – si basano su architetture di machine learning in grado di elaborare testi complessi, umanizzare conversazioni e generare contenuti in linguaggio naturale quasi indistinguibili da quelli umani³.

In virtù di queste caratteristiche tecniche le aree applicative dei modelli computazionali sono ormai sempre più estese: customer care, sanità digitale, istruzione, giornalismo, comunicazione istituzionale e politica, consulenza legale e finanziaria, marketing e commercio.

Se il loro funzionamento era in principio fondato su un addestramento proveniente dall'uomo e finalizzato a riconoscere, scegliere e decidere in modo autonomo, oggi le macchine dotate di capacità di apprendimento automatico sono in grado di addestrarsi da sole, imparando dalle proprie decisioni e dai dati che raccolgono autonomamente, e ciò porta a una intrinseca impossibilità di prevedere le scelte fatte dall'algoritmo e a una intrinseca opacità dello stesso (che si definisce perciò black box).

Ne deriva perciò un cambio di paradigma. I chatbot rischiano di non essere considerati esclusivamente come meri strumenti di supporto alle decisioni umane, ma alla stregua di intermediari cognitivi che diventano mezzo per l'accesso all'informazione capaci di influenzare le decisioni individuali e collettive⁴.

Oggi, invece, assistiamo a una trasformazione profonda e ulteriore: i modelli più evoluti integrano sistemi di *affective computing* o *emotion recognition*, in grado di dedurre lo stato emotivo dell'utente analizzando caratteristiche fisiche, fisiologiche e segnali non verbali come le variazioni nel tono della voce, le microespressioni facciali, le scelte lessicali e i tempi di risposta. Il loro funzionamento è basato su un'inferenza statistica assunta tra determinati

La conclusione di Searle è che la sintassi non coincide con la semantica, ovvero che la semplice elaborazione dei simboli (come fanno le macchine) non è sufficiente a produrre una comprensione o una coscienza autentiche. Pertanto, eseguire un compito come fa una macchina non implica che quest'ultima capisca o esprima una coscienza.

³ E. LORU, J. NUDO, N. DI MARCO, A. SANTIROCCHI, R. ATZENI, M. CINELLI, V. CESTARI, C. ROSSI-ARNAUD, W. QUATTROCIOCCI, *The simulation of judgment in LLMs*, in *Proc. Natl. Acad. Sci. U.S.A.*, 2025, 122, 42, e2518443122. Consultabile al sito <https://doi.org/10.1073/pnas.2518443122>.

L'intelligenza artificiale può ottenere gli stessi risultati di un essere umano, ma il percorso attraverso il quale li raggiunge non è il medesimo, poiché la macchina opera attraverso approssimazioni linguistiche e significati predefiniti. Non è in grado di cogliere le sfumature che guidano il giudizio come può fare un essere umano. Questo fenomeno è chiamato «epistemia»: una condizione in cui l'apparenza di un giudizio coerente e autorevole deriva esclusivamente da modelli statistici, producendo l'illusione della conoscenza sebbene in realtà la verosimiglianza superficiale sostituisca il ragionamento basato sulle evidenze.

⁴ L. FLORIDI, *The ethics of information*, Oxford, 2013; A. MCSTAY, *Emotional AI, soft biometrics and the surveillance of emotional life: An unusual consensus on privacy*, in *Big Data Soc.*, 2020, 7, 1, 1–17. Consultabile su <https://doi.org/10.1177/2053951720904386>. Questi sistemi esprimono il potenziale per erodere anche la capacità di autodefinizione di sé stessi, cagionando un'enorme compromissione dell'identità della persona umana

tratti biologici, gli stati emotivi e la loro successiva categorizzazione, che consente di trarre informazioni attraverso il confronto tra il tratto rilevato e il set di dati di riferimento⁵.

Questi sistemi, che promanano da ricerche nell'ambito delle neuroscienze⁶, hanno l'obiettivo di individuare lo stato emotivo dell'interlocutore, pertanto, non si limitano a generare risposte testualmente pertinenti, ma adattano il contenuto e il registro comunicativo secondo le presunte emozioni dell'utente.

Le applicazioni pratiche possono essere molteplici e collocarsi in settori anche molto diversi tra loro: ad esempio, un chatbot che inferisce emozioni nell'ambito sanitario potrà modulare le proprie risposte di fronte a segnali percepiti come di ansia o di paura; nel contesto commerciale potrà intensificare strategie persuasive se rileva esitazione nella voce del cliente⁸.

Occorre soffermarsi in particolare sul contesto dell'istruzione scolastica poiché il dibattito sopra l'utilizzo di tali sistemi è fervente nelle scienze della formazione⁹. Gli odierni sistemi di apprendimento adattivo sono ambienti tecnologici facilitatori che personalizzano contenuti, modalità e ritmi di studio sulla base delle caratteristiche, preferenze e conoscenze pregresse dello studente; aiutano gli insegnanti a intervenire in modo mirato e migliorano l'efficacia dell'apprendimento. I sistemi di intelligenza artificiale, opportunamente addestrati, possono formulare inferenze probabilistiche sull'andamento della formazione e "raccomandare" o suggerire a chi apprende nuovi contenuti o strategie di apprendimento¹⁰.

⁵ L. STARK, J. HOEY, *The Ethics of Emotion in Artificial Intelligence Systems*, in *FAccT 21: 2021 ACM Conference on Fairness, Accountability, and Transparency*, 2021, 782-783, doi: 10.1145/3442188.3445939.

⁶ Si tratta di studi orientati a individuare come le emozioni umane incidano sul comportamento dell'uomo stesso, cfr. J. S. LERNER, Y. LI, P. VALDESOLO, K. S. KASSAM, *Emotion and Decision-Making*, in *Annu. Rev. Psychol.*, 2015, 66, 799-823.

⁸ È stato oggetto di studio il meccanismo manipolatorio attraverso il quale i chatbot realizzano un'interazione con l'utente finalizzata a prolungare il tempo di permanenza all'interno della conversazione, ad esempio. Cfr. J. DE FREITAS, Z. OGUZ-UGURALP, A. KAAAN-UGURALP, *Emotional Manipulation by AI Companions*, arXiv e-prints, no. arXiv:2508.19258, 2025, doi:10.48550/arXiv.2508.19258.

⁹ Il primo ideatore delle machine di apprendimento in ambito educativo è stato Skinner, il padre della teoria del condizionamento operante, cfr. B. F. SKINNER, *The science of learning and the art of teaching*, in *Harvard educ. Rev.*, 1954, 29, 86-97; Id., *Verbal behavior*, Appleton-Century-Crofts, New York, 1957; Id., *Teaching machines*, in *Science*, 1958, 128, 969-977.

Egli diede vita alle cosiddette teaching machines (la più rudimentale, Glider, era una scatola che presentava una finestra interna attivabile tramite una leva che manifestava quattro possibili risposte immediate a domande poste da uno studente che aveva la possibilità di controbattere alle prime risposte): avevano lo scopo di stimolare i comportamenti desiderati tramite un percorso rigoroso e sequenziale di apprendimento, controllando l'attenzione dello studente con contenuti lineari identici per tutti (si trattava di un metodo ispirato all'esperimento della Skinner box, in cui un organismo compie delle azioni in base ai rinforzi - stimoli positivi o negativi - che riceve), permettevano ai discenti di studiare al proprio ritmo, aumentando la motivazione tramite meccanismi di riscontro immediato e liberando l'insegnante da una parte della gestione della valutazione. Cfr., U. ZONA, F. BOCCI, *La rete come una skinner box. neocomportamentismo, bolle sociali e post-verità*, in *Media Education. Studi, ricerche, buone pratiche*, 2018, 9, I, 57-77 doi: 10.14605/MED911804; R. HORN, J. KINCHELOE, *The Praeger Handbook of Education and Psychology*, Bloomsbury Publishing USA, Westport Connecticut, Londra, 2006, 1, 201-205; N. KARA, N. SEVIM, *Adaptive Learning Systems: Beyond Teaching Machines*, in *Contemporary Educational Technology*, 2013, 4, II, 108-120. Di contro questi strumenti non favorivano l'interazione sociale, ignoravano i bisogni emotivi degli studenti e imponendo un percorso obbligato, causavano demotivazione. Cfr., J. G. HOLLAND, *Teaching machines: an application of principles from the laboratory*, in *Journal of the Experimental Analysis of Behavior*, 1960, 3, IV, 275-287.

¹⁰ V. FRANZONI, G. BIONDI, D. PERRI, O. GERVASI, *Enhancing Mouth-Based Emotion Recognition Using Transfer Learning*, in *Sensors* 20, 2020, 18, 5222. Consultabile su <https://doi.org/10.3390/s20185222>.

L'avanzamento tecnico di tali sistemi, del tutto (f)attuale, fa emergere rilevanti questioni giuridiche che in questo elaborato verranno analizzate, onde rifuggire il pericolo di interferenze invasive nella sfera privata individuale e nell'esercizio dei diritti fondamentali dell'uomo, dal momento che spetta al giurista rivendicare il ruolo attivo nella modellazione giuridica dei fenomeni tecnologici, nell'ottica del[...] «la creazione di valore [che] non deriva dall'hype o dalla speculazione, ma dallo sviluppo di tecnologie che risolvono problemi del mondo reale e migliorano la vita delle persone e gli ambienti»¹².

2. Quando l'etica si intreccia con il diritto. Domande di senso e questioni giuridicamente rilevanti.

Se si riflette sull'evoluzione tecnologica, da qualunque latitudine la si guardi, facilita la riflessione adottare una prospettiva interdisciplinare¹³. È così che si pone la questione, ancor prima etica che giuridica, di evitare che il trattamento di dati "emotivi" riduca la persona a mero oggetto di misurazione e veicolo commerciale di persuasione. In particolare, emerge il rischio causato dalla "datafication": l'operazione di creazione e memorizzazione di un modello unico che identifica gli individui sulla base di una serie di dati univoci. Una volta che il modello è stato creato e memorizzato, chiunque ne sia in possesso può tracciare e riconoscere quell'individuo ovunque nel mondo e potenzialmente per qualsiasi scopo. Risulta evidente che poiché tali caratteristiche identificative sono inevitabilmente parte della persona, la loro raccolta e uso interferisce con l'autonomia personale e la dignità dell'uomo, anche in ragione della particolare difficoltà da parte delle persone fisiche di essere realmente consapevoli che un trattamento dei loro dati è in corso e della conseguente difficoltà a sottrarsi. In tal senso la capacità di un algoritmo di realizzare una individuazione o predizione sullo stato emotivo di un individuo, tanto da categorizzarlo, potrebbe mettere a rischio anche il controllo delle persone fisiche sui propri stati emozionali (*rectius*, su come questi vengono identificati)¹⁴.

In quest'ottica occorre superare una visione ingenuamente positiva dell'innovazione, riconoscendo che lo sviluppo tecnologico necessita di essere guidato da scelte consapevoli e moralmente fondate. Progettare un sistema in grado di riconoscere e interpretare le emozioni implica altresì l'assunzione della responsabilità che queste siano rilevate in maniera attendibile e di non

¹² L. FLORIDI, *Why the AI Hype is Another Tech Bubble*, 2024. Consultabile su <http://dx.doi.org/10.2139/ssrn.4960826>. Prosegue l'Autore affermando «La sfida sta nel distinguere tra il potenziale genuino dell'AI e le sue promesse ridicole, e nel canalizzare investimenti e sforzi in aree che produrranno benefici sostenibili a lungo termine per la società. Perché la bolla AI scoppi delicatamente e il prossimo inverno AI sia mite, è meglio agire ora».

¹³ L. FLORIDI, et al., *AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, in *Minds & Machines*, 2018, 28, IV, 689–707. Consultabile su <https://doi.org/10.1007/s11023-018-9482-5>; Id., *Soft Ethics and the Governance of the Digital*, in *Philos. Technol.*, 2018, 31, I, 1–8. Consultabile su <https://doi.org/10.1007/s13347-018-0303-9>.

¹⁴ R. S. KIM, A. MOREAU, B. CHEVALIER, et al., *Affective Sovereignty: A Minimal Declaration on Emotional Interpretation Rights in the Age of Algorithmic Power: When AI Interprets My Emotions Before I Do, Can I Still Call Them My Own?*, PREPRINT *Research Square*, 2025. Consultabile su [<https://doi.org/10.21203/rs.3.rs-7157255/v1>].

manipolarle¹⁵. Tale principio dovrebbe operare su due livelli distinti ma interconnessi: un livello etico, volto ad accrescere la consapevolezza riguardo alle implicazioni morali insite nella tecnologia; e un livello politico, finalizzato a orientare lo sviluppo dell'innovazione verso decisioni informate da valori condivisi e da una responsabilità collettiva¹⁶.

Il giurista non può prescindere dall'adottare un punto di vista di tutela dei diritti fondamentali dell'uomo, con particolare riferimento agli art. 7 e 8 della Carta dei diritti fondamentali dell'UE, i quali proteggono rispettivamente la vita privata¹⁷ e i dati personali degli individui¹⁸. Per ragioni di delimitazione della trattazione verranno taciute le prospettive di comparazione con gli ordinamenti extra-UE, che in realtà rappresentano una fonte di approfondimento rilevante per uno studio provocatorio del tema¹⁹.

Adottando la prospettiva della tutela dei diritti fondamentali si può osservare come i rischi per una tecnologia che sfrutta le caratteristiche biologiche dell'individuo a fini di identificazione e/ o autenticazione possono

¹⁵ Le «Ethics guidelines for trustworthy AI» dell'8 aprile 2019 si soffermano sul principio di spiegabilità algoritmica e sull'importanza della trasparenza di una IA affidabile e umano-centrica. Tali linee guida delineano un'intelligenza artificiale meritevole di fiducia o affidabile, che per essere tale deve rispettare le norme e la legalità, osservare i principi etici e mostrare robustezza. A tali fini, l'«IA affidabile» deve rispondere e deve essere valutata sulla base di quattro principi etici (rispetto dell'autonomia umana; prevenzione dei danni; equità; spiegabilità) e di sette requisiti fondamentali (intervento e sorveglianza umani; robustezza tecnica e sicurezza; riservatezza e governance dei dati; trasparenza; diversità, non discriminazione ed equità; benessere sociale ed ambientale; accountability).

¹⁶ G. TAMBURRINI, *Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiale*, Roma, 2020. Questa prospettiva riflette la necessità di un approccio etico nel progresso tecnologico che eviti semplicistiche narrazioni di progresso e preveda invece meccanismi di governance atti a garantire responsabilità, trasparenza e sostenibilità, elementi fondamentali per un'innovazione tecnologica al servizio della società e nel rispetto dei diritti umani. I dilemmi morali sollevati dall'integrazione di capacità emotive nelle tecnologie richiedono quindi non solo riflessioni teoriche, ma anche azioni concrete che abbraccino la dimensione politica e sociale dello sviluppo tecnologico, promuovendo un uso equo, inclusivo e sostenibile delle tecnologie emergenti.

¹⁷ In continuità con l'art. 8 della Convenzione Europea dei Diritti dell'Uomo che sancisce il diritto alla protezione della vita privata, la quale comprende l'integrità dell'intimità della persona fisica e la sua identità. In proposito è importante sottolineare le osservazioni dell'Illustre Maestro C. M. BIANCA, il quale evidenziava «il diritto alla riservatezza come diritto alla non divulgazione dei fatti personali è ormai confluito nel diritto alla protezione dei dati personali, perdendo autonoma rilevanza. Autonoma rilevanza conserva invece il diritto alla riservatezza quale diritto della persona al rispetto della propria vita privata», cfr., C. M. BIANCA, F. D. BUSNELLI (a cura di), *La Protezione dei dati personali – Commentario al D. lgs. 30 giugno 2003, n. 196 («Codice della privacy»)*, Padova, 2007, I, 21.

¹⁸ Ai due Illustri Autori di cui *supra* si deve una qualificazione *ante litteram* del diritto alla protezione dei dati personali come un diritto autonomo della personalità, dotato di «un proprio contenuto e una propria disciplina» e per questo «sanzionat[o] a prescindere dall'accertamento della violazione di altri diritti», cfr. C. M. BIANCA, F. D. BUSNELLI, *op. cit.*, 21; a dispetto di una dottrina che spaziava da un'interpretazione come diritto «secondario» della protezione dei dati personali o che proponeva il paradigma «reificante» del dato personale.

¹⁹ L'analisi comparata della disciplina richiederebbe un metodo e una trattazione apposita, in questa sede mi limito a citare alcuni primi e certamente non esaurienti riferimenti. Numerosi fornitori hanno sede extra-UE, e i dati emotivi seguono i medesimi flussi degli altri dati personali, con il rischio di uscire dall'ombrello protettivo europeo. Ricordiamo in proposito l'approccio più flessibile degli Stati Uniti, che privilegia linee guida settoriali e autoregolamentazione, e quello più prescrittivo della Cina, con un forte controllo centralizzato sull'uso dei dati e sulla conformità dei contenuti, cfr., in particolare, «Avviso della Cyberspace Administration of China relativo alla consultazione pubblica sulla «Misura per l'identificazione dei contenuti sintetici generati dall'intelligenza artificiale (bozza per consultazione pubblica)» del 1 settembre 2025, tramite cui l'amministrazione cinese dispone che per ogni contenuto generato dall'IA provider, piattaforme online, app store e utenti siano vincolati a identificare chiaramente tutti i contenuti prodotti da IA, indipendentemente dal formato (testo, audio, video, immagini, realtà virtuale). In un'ottica di politica regolatoria appare necessario favorire convergenze normative fra Europa, Stati Uniti, Cina e altre giurisdizioni onde evitare pratiche di forum shopping regolatorio e abbassamento degli standard di tutela.

essere molteplici²⁰. A titolo meramente esemplificativo si possono elencare alcuni dei più evidenti rischi derivanti dal trattamento dei dati da parte di questi modelli. In primo luogo, ciò che è evidente: l'individuazione dell'emozione è frutto di un raffronto probabilistico, con tutto ciò che questo meccanismo di funzionamento comporta in termini di attendibilità e affidabilità delle determinazioni algoritmiche. Si deve in tal senso evitare di alimentare forme di profilazione illecita e discriminazioni algoritmiche (si pensi, tra le altre, alla raccolta e profilazione dei dati etnici o religiosi che porti a discriminazioni ingiustificate sotto il profilo della repressione e prevenzione penale). I dati dedotti dalle emozioni non si limitano a descrivere, ma possono essere direttamente impiegati per influenzare la libertà decisionale e addirittura contrattuale dell'individuo, fino a scandire una sorta di effetto manipolativo indiretto.

Inoltre, molti sistemi di IA mantengono memorie storiche delle conversazioni a fini di autoapprendimento, occorre in tal senso domandarsi se e come la persona possa esercitare diritti di accesso, rettifica e cancellazione dei dati stessi. Infine, la complessità dei sistemi IA rende difficile una reale comprensione da parte degli utenti delle modalità e finalità di raccolta e impiego dei dati emotivi.

Sul piano tecnico il trattamento di dati da parte di chatbot dotati di intelligenza artificiale e capaci di autoapprendimento implica due dimensioni strettamente collegate: l'addestramento dei modelli e il loro utilizzo nell'interazione uomo-macchina.

I modelli vengono addestrati su dataset di enormi dimensioni e quando l'utente conversa con un chatbot, può condividere, intenzionalmente o meno, dati personali o particolari (talora relativi alle proprie credenze più intime, si pensi ad esempio allo studente di medicina che si rivolga al chatbot per avere consigli su quale specializzazione intraprendere e condivida in chat la propria adesione valoriale ad esempio all'"astenersi dal sangue" per credo religioso, e poi scelga una specializzazione come medicina interna o chirurgia generale).

Non solo, un cittadino che chiede a un assistente virtuale consigli sanitari sta fornendo implicitamente informazioni sulla propria salute.²¹ Analogamente, uno studente che carica materiali didattici personali condivide dati intellettuali e potenzialmente identificativi.

La successiva elaborazione e memorizzazione di tali dati da parte del modello rischia di creare forme di sorveglianza implicita o di riutilizzo del dato non lecito, per tacere del sempre presente rischio di "data breach".

È evidente che per perseguire la finalità del trattamento non parliamo più solo di dati personali in senso stretto²² – nome, indirizzo, contatti – ma anche di

²⁰ C. WENDEHORST, Y. DULLER, *Biometric Recognition and Behavioural Detection*, in *Policy Department for Citizens' Rights and Constitutional Affairs EN Directorate-General for Internal Policies*, 2021, 12. Consultabile su [europarl.europa.eu/RegData/etudes/STUD/2021/696968/IPOLSTU\(2021\)696968_EN.pdf](http://europarl.europa.eu/RegData/etudes/STUD/2021/696968/IPOLSTU(2021)696968_EN.pdf).

²¹ A. BERTOLINI, G. AIELLO, *Robot companions: A legal and ethical analysis*, in *The Information Society*, 2018, 34, III, 130-140, DOI: 10.1080/01972243.2018.1444249.

²² Ai sensi dell'art. 4, par. 1, punto 1) del GDPR si qualifica «[...] «dato personale»: qualsiasi informazione riguardante una persona fisica identificata o identificabile («interessato»); si considera identificabile la persona fisica che può essere identificata, direttamente o indirettamente, con particolare riferimento a un identificativo come il nome, un numero di identificazione, dati relativi all'ubicazione, un identificativo online o a uno o più elementi caratteristici della sua identità fisica, fisiologica, genetica, psichica, economica, culturale o sociale; [...]».

dati “particolarissimi” (*caveat*, questo termine è da intendersi in senso meramente divulgativo): i dati biometrici e tratti comportamentali ed emotivi che sono capaci di incidere direttamente sull’identità e sull’autodeterminazione della persona. Queste informazioni dal punto di vista della macchina costituiscono la base per modificare e adattare il comportamento del sistema, con l’obiettivo di migliorare l’esperienza dell’utente, incidere sulla sua soddisfazione e, in determinati contesti, finanche influenzarne le decisioni e le azioni.

I rischi per i diritti fondamentali da affrontare in questo contesto sono principalmente collegati alla rilevazione dell’emozione e al modo in cui viene utilizzata per tracciare un profilo (portando, ad esempio, allo sfruttamento di vulnerabilità identificate²⁴)²⁵.

Avendo chiarito il perimetro dell’analisi, si può ora procedere all’approfondimento della disciplina di riferimento. Per definire i dati biometrici il punto di partenza non può che essere l’articolo l’art. 4, paragrafo 14, del GDPR, per il quale questi sono «dati personali ottenuti da un trattamento tecnico specifico relativo a caratteristiche fisiche, fisiologiche o comportamentali di una persona fisica, che consentono o confermano l’identificazione univoca»²⁶.

Emerge dunque come il dato biometrico debba far riferimento a una persona identificata, e non semplicemente “identificabile”, perché è proprio l’operazione tecnica specifica con finalità di estrazione dei dati biometrici e registrazione degli stessi che rende idonei questi dati alla nuova identificazione della persona in operazioni successive²⁷. Il rilevamento delle emozioni, pur non sempre coincidente con l’identificazione univoca del soggetto²⁸, si colloca ai confini di questa categoria: quando deriva da analisi biometriche, come il monitoraggio della voce o del volto, i dati trattati rientrano nelle categorie

²⁴ Ad esempio, la persona potrebbe essere associata ad un determinato tratto psicologico che la rende “inidonea” ad un ruolo, oppure, valutata la sua propensione al consumo e alla spesa, venirle proposto un prezzo per uno stesso bene di consumo molto più alto a parità di condizioni rispetto ad altri soggetti. A tal proposito si vedano A. BERTOLINI, G. AIELLO, *op. cit.*, ibidem. Gli A. evidenziano come proprio i soggetti più vulnerabili, con una particolare attenzione agli individui minori di età, rappresentino il target più esposto a manipolazioni derivanti dall’interazione tra uomo e macchina.

²⁵ Non meno rilevante è la questione del profiling automatizzato, disciplinato dall’articolo 22 GDPR. Qualora il chatbot utilizzi il riconoscimento delle emozioni e profili il soggetto determinando conseguenze “significative” che impattano sulla vita dell’utente – per esempio negare l’accesso a un servizio sanitario, graduare un’offerta economica, modulare livelli di assistenza, valutazioni di idoneità o concessione di servizi – esso cade sotto le prescrizioni del GDPR che lo vietano al fine di preservare il rispetto dei diritti fondamentali dell’individuo.

Il divieto opera salvo che non si sia in presenza di rigorose garanzie, come il diritto ad esigere un intervento umano, la possibilità di esprimere il proprio punto di vista, la tutela contro la discriminazione e l’arbitrarietà. Si parla di conseguenze significative da intendersi come lesioni dei diritti fondamentali o degli interessi del soggetto interessato dal trattamento, da valutare caso per caso. Cfr., Corte di Giustizia dell’Unione Europea, 7 dicembre 2023, OQ c. Land Hessen (ECLI:EU:C:2023:957). Consultabile su eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62021CJ0634.

²⁶ Conforme al considerando 51 del GDPR.

²⁷ In effetti, dall’art. 4. par. 14 del GDPR si evince come la definizione di dato biometrico sia volta a delimitare il campo di applicazione della norma a quei dati “univoci”; quindi, immutabili nel tempo in quanto connaturati all’essenza della persona stessa e volta ad escludere quei dati che, pur in astratto fonte idonea di informazioni biometriche, non sono sottoposti a un trattamento tecnico idoneo allo scopo. Cfr., R. MONTINARO, *Emotion Recognition and Personalized Advertising*, in *European Review of Private Law*, 2024, VI, 1003–1036, fa riferimento in tal senso, all’esclusione dei cd. “raw data”.

²⁸ Si ricade, in questi casi, sotto l’ombrello dell’art. 4 par. 1, nei casi in cui l’identificazione univoca non sia stata effettuata.

particolari di cui all'art. 9 GDPR²⁹. Tali dati sono soggetti a un regime di protezione rafforzato, il cui trattamento è generalmente vietato salvo specifiche deroghe, quali: il consenso esplicito dell'interessato; motivi di interesse pubblico rilevante; tutela della vita e salute della persona; esercizio di diritti legali o giudiziari o esecuzione di obblighi contrattuali; l'interessato ha reso manifestamente pubblici questi dati. Perché ciò avvenga, è opportuno ribadire come la finalità esplicita del trattamento dei dati debba essere quella dell'identificazione biometrica; se infatti la finalità fosse altra allora la categoria dei dati trattati non sarebbe "particolare" ai sensi dell'art. 9³⁰.

In presenza di trattamento di dati biometrici l'attenzione del titolare dovrà essere orientata sull'attuazione dei principi di necessità e proporzionalità, valutando le condizioni del caso concreto al fine di valutare qualunque opzione possa essere meno invasiva al fine di erogare il servizio, in proposito è particolarmente attuale il parere dell'EDPB n. 11 del 23 maggio 2024 sull'uso del riconoscimento facciale per snellire il flusso dei passeggeri negli aeroporti³⁶.

L'EDPB individua quattro scenari, di cui due considerati accettabili rispetto ai principi di necessità e proporzionalità e due troppo invasivi. In particolare, soluzioni che privilegiano la riservatezza dell'utente come «la conservazione di un modello biometrico registrato nelle mani della persona [...]» o «la conservazione centralizzata all'interno dell'aeroporto di un modello biometrico registrato in forma cifrata con una chiave segreta nota esclusivamente al passeggero [...]».

Al contrario, appare raggiungibile con forme meno invasive la finalità del trattamento ove ricorra «la conservazione centralizzata di un modello biometrico registrato in forma cifrata all'interno dell'aeroporto sotto il controllo del gestore aeroportuale [...] oppure la conservazione centralizzata di un modello biometrico registrato in forma cifrata nel cloud sotto il controllo della compagnia aerea o del suo fornitore di servizi cloud [...]». Di fatto, «dal momento che la conservazione dei dati identificativi e biometrici avviene in una banca dati centrale, se la riservatezza di quest'ultima è compromessa, ciò può successivamente comportare l'accesso all'intero insieme di dati e consentire l'identificazione non autorizzata o illecita dei passeggeri in altri ambienti»³⁷.

²⁹ Questi sono: «[...] l'origine razziale o etnica, le opinioni politiche, le convinzioni religiose o filosofiche, o l'appartenenza sindacale, nonché [...] dati genetici, dati biometrici intesi a identificare in modo univoco una persona fisica, dati relativi alla salute o alla vita sessuale o all'orientamento sessuale della persona».

³⁰ European Data Protection Board Guidelines, parere n. 3/2019 sul trattamento dei dati personali per mezzo di apparati di videoregistrazione, par. 5, punto 80. Consultabile su edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201903_video_devices_en_0.pdf.

³⁶ L'EDBP si è espresso su istanza del Garante Privacy francese riguardo la compatibilità con l'articolo 5, paragrafo 1, lettere e) e f), e gli articoli 25 e 32 del GDPR. È importante sottolineare che non viene analizzato, dal momento che l'EDBP si esprime nei limiti dell'istanza presentatagli, il punto riguardante la base giuridica del trattamento. Peraltro, il tema è particolarmente rilevante in termini di identificazione di categorie particolari di dati, in quanto viene diversificata la base giuridica del trattamento ai sensi dell'art. 6 GDPR, per approfondire R. MONTINARO, *op. cit.*, ibidem.

³⁷ Ricorrendo il primo degli scenari considerati incompatibili con le regole europee in materia di data protection, l'11 settembre scorso il Garante della Privacy, conformemente al citato parere, ha sospeso l'adozione della specifica soluzione tecnologica adottata e dunque l'erogazione del servizio di cd. faceboarding dell'Aeroporto di Linate (società SEA Aeroporti), in quanto incompatibile con la vigente disciplina europea sulla protezione dei dati personali.

Non sempre il riconoscimento delle emozioni si fonda sul trattamento di dati biometrici tout court. Un primo elemento di distinguo è rappresentato, come già anticipato, dallo scopo del trattamento dei dati, che non necessariamente consiste nella identificazione biometrica in quanto tale (basti pensare all'utilizzo dei dati allo scopo di personalizzazione dei percorsi formativi³⁸, o ancora alle attività di categorizzazione a scopo promozionale, commerciale e pubblicitario). In seconda istanza, il trattamento tecnico dei dati da parte di questi sistemi non necessariamente si basa su dati biometrici in senso stretto, ma anzi spesso individua elementi non così "sensibili" appartenenti a un soggetto, di natura più propriamente comportamentale (come i lineamenti del volto, la risata, le variazioni del tono della voce) per distribuire e segmentare le persone in base a una categorizzazione già individuata³⁹: il trattamento si concretizza così in una forma di profilazione⁴⁰. In realtà ciò non esclude che i dati siano comunque qualificabili come personali ai sensi dell'art. 4 par. 1, in ossequio alla dottrina Nowak della CGUE⁴¹, la quale sostiene che le informazioni che sono qualificabili come dato personale sono «non solo oggettive ma anche soggettive, sotto forma di opinioni e valutazioni, a condizione che siano "relative" all'interessato». Del resto, attraverso questa forma di profilazione si ottengono risultati orientati sulla persona e che vengono utilizzati allo scopo erogare un servizio targetizzato sulla base delle caratteristiche dedotte della persona in questione⁴².

A livello regolatorio, perciò, i principi chiave del GDPR (come descritti dall'art. 5)⁴³ restano applicati rigorosamente⁴⁴ anche nella gestione dei tratti

³⁸ Un utilizzo del tutto compliant alla normativa vigente, per esempio, è quello di consentire attraverso i cd. *learning analytics* l'analisi dei risultati di apprendimento di una popolazione di studenti, e identificare tendenze e pattern di comportamento che permettono di individuare con una buona probabilità le classi di studenti a rischio di abbandono durante un percorso formativo, in questo modo procedendo a individuare interventi mirati a ridurre il rischio stesso, cfr. R. BAKER, G. SIEMENS, *Educational data mining and learning analytics*, in *The Cambridge Handbook of the Learning Sciences*, Cambridge Handbooks in Psychology, R. K. SAWYER (a cura di), Cambridge, 2014, 253-272; I. S. FITTON, D. J. FINNEGAN, M. J. PROULX, *Immersive virtual environments and embodied agents for e-learning applications*, in *PeerJ Comput Sci*, 2020, VI, e315.

³⁹ Cfr., documento del Working Group art. 29 sulla biometria (2003) WP 80 e parere n. 3/2012 sugli sviluppi delle tecnologie biometriche (2012), WP 193. Tali elaborati sono antecedenti all'emanazione del GDPR, ma rimangono attuali anche con la sua entrata in vigore.

⁴⁰ M. HILDEBRANDT, S. GUTWIRTH, *Profiling the European Citizen Cross-Disciplinary Perspectives*, Springer, 2008, 241.

⁴¹ Corte di Giustizia dell'Unione europea, 20 dicembre 2017, *Peter Nowak contro Data Protection Commissioner*, Causa C-434/16, punto 34, curia.europa.eu/juris/liste.jsf?num=C-434/16; Corte di giustizia dell'Unione europea, 7 marzo 2024, *IAB Europe contro Gegevensbeschermingsautoriteit*, punto 37, curia.europa.eu/juris/liste.jsf?num=C-604/22.

⁴² V. RUPP, M. VONGRAFENSTEIN, *Clarifying "personal data" and the role of anonymization in data protection law: including and excluding data from the scope of the GDPR (more clearly) through refining the concept of data protection*, in *Computer Law & Security Review*, 2024, LII, 4.

⁴³ Liceità, correttezza e trasparenza (art. 5, par. 1, lett. a): ogni trattamento deve avere una base giuridica valida, essere corretto e trasparente per l'interessato, che deve essere informato in modo chiaro e completo circa le finalità e modalità trattamentali; limitazione della finalità (art. 5, par. 1, lett. b): i dati devono essere raccolti per finalità determinate, esplicite e legittime, e non trattati in modo incompatibile con tali finalità; minimizzazione dei dati (art. 5, par. 1, lett. c): è necessario limitare la raccolta e l'elaborazione solo a dati strettamente necessari allo scopo; precisione (art. 5, par. 1, lett. d): i dati devono essere esatti, aggiornati e sono richiesti meccanismi di rettifica e cancellazione; limitazione della conservazione (art. 5, par. 1, lett. e): i dati devono essere eliminati o resi anonimi una volta esaurite le finalità; integrità e riservatezza (art. 5, par. 1, lett. f): devono essere impiegate misure tecniche e organizzative adeguate a proteggere i dati dalle violazioni.

⁴⁴ Si veda la vicenda intercorsa tra Garante della privacy e Luka Inc., società produttrice del chatbot Replika. Al provvedimento di limitazione provvisoria del trattamento dei dati personali iniziale, ha fatto seguito una

emotivi non derivanti dal trattamento di dati biometrici (e dunque quando non entrano in gioco le valutazioni descritte dall'art. 9 del GDPR)⁴⁵, sebbene si sia sostenuto che queste misure non siano sufficientemente protettive per la tutela della riservatezza della persona⁴⁶.

3. AI Act & GDPR: una coppia che “funziona”

L'implementazione dell'AI Act rappresenta una novità importante, provvedendo in particolare sulla profilazione emotiva. Il Reg. UE n. 2024/1689⁽⁴⁷⁾ si pone in una prospettiva di continuità e completamento rispetto alla cornice protettiva fornita dal GDPR. Invero, viene disegnata una classificazione dei sistemi di intelligenza artificiale secondo un approccio basato sul rischio (la piramide del rischio), che può essere evidenziato dal divieto di immissione sul mercato di determinati sistemi (ad esempio i cd. deepfake) alla gestione dei fattori di rischio nelle applicazioni ad alto rischio (ad esempio le auto a guida autonoma o la robotica sanitaria).

Il Regolamento è uno strumento normativo ispirato dalla logica della neutralità tecnologica e votato al principio di trasparenza e spiegabilità algoritmica (dei processi di addestramento e del linguaggio), che nasce con lo scopo di creare un mercato unico per l'Intelligenza Artificiale, prevenendo la frammentazione del mercato e fornendo ai suoi attori la certezza giuridica di cui hanno bisogno per operare in tutti gli stati membri dell'UE. D'altra parte, e tale aspetto è essenziale in questa trattazione, si tratta di uno strumento caratterizzato da valori che limitano l'accesso al mercato unico proprio ai sistemi di IA “affidabili”, in particolare in ossequio alla tutela dei diritti fondamentali, con lo scopo di assicurare il controllo umano e la protezione dei dati personali per l'uso dei sistemi di intelligenza artificiale⁴⁸.

L'AI Act segna una particolare attenzione ai sistemi che utilizzano dati “biometrici” con l'obiettivo di limitarne l'uso in contesti particolarmente vulnerabili e di garantire adeguati livelli di protezione onde evitare compromissioni dei diritti fondamentali della persona⁴⁹.

vera e propria sanzione pecuniaria derivante dalla mancata adozione di misure idonee alla tutela dei minori, cfr. GPDP, Provvedimento del 10 aprile 2025 [doc.web 10130115], Consultabile al sito www.garanteprivacy.it e per un commento più approfondito A. DI CERBO, *Provocazioni culturali e implicazioni giuridiche degli AI Companion*, in *EJPLT*, 2025, II.

⁴⁵ R. MONTINARO, *op. cit.*, ibidem.

⁴⁶ E. STEINDL, *Does the European Data Protection Framework Adequately Protect Our Emotions? Emotion Tech in Light of the Draft AI Act and Its Interplay with the GDPR*, in *European Data Protection Law Review*, 2022, VIII, 311.

⁴⁷ Nel proseguo della trattazione verrà citato come il “Regolamento”.

⁴⁸ Tali principi sono fatti propri anche dalla legge italiana di regolazione dell'intelligenza artificiale, la prima forma di regolazione espressa da un ordinamento interno all'UE capace di integrare le prescrizioni del Regolamento 2024/1689. Cfr. L. n. 132/2025.

⁴⁹ Sebbene nelle fasi preparatorie degli Autori auspicassero una disciplina legale più restrittiva in materia di certificazioni di conformità, FRIA e obblighi di trasparenza, cfr. I. BARKANE, *Questioning the EU proposal for an Artificial Intelligence Act: The need for prohibitions and a stricter approach to biometric surveillance*, in *Information Polity*, 2002, vol. 1, XXVII, 2, 147-162. Consultabile su <https://doi.org/10.3233/IP-211524>. In particolare, la FRIA (Fundamental Rights' Impact Assessment) non era prevista nella prima bozza del Regolamento, ed ora si colloca (art. 27 AI Act) in continuità e completamento con la DPIA già prevista ai sensi dell'art. 35 del GDPR. Di fatto, sono entrambi strumenti che condividono lo scopo di individuare preventivamente e mitigare danni potenziali, ciò che li differenzia è il terreno di riferimento, in quanto la

In questo schema, i sistemi di riconoscimento delle emozioni in contesti lavorativi, educativi o sanitari sono stati qualificati ad “alto rischio”, con obblighi di *compliance* rafforzati ⁵⁰. In alcune declinazioni – come nel caso del riconoscimento emotivo applicato agli ambienti scolastici per monitorare l'attenzione degli studenti – il legislatore europeo ha introdotto un vero e proprio divieto⁵¹.

In particolare, l'articolo 3, par. 39 definisce i sistemi di riconoscimento delle emozioni come sistemi di IA «finalizzati all'identificazione o alla desunzione delle emozioni o delle intenzioni delle persone fisiche sulla base dei loro dati biometrici». In combinato con il considerando 18 del Regolamento possiamo qualificare le emozioni o intenzioni, che sono «felicità, tristezza, rabbia, sorpresa, disgusto, imbarazzo, eccitazione, vergogna, disprezzo, soddisfazione e divertimento»⁵²; la definizione data non solo è in continuità con quella già vista nel GDPR, ma addirittura amplia il perimetro soggettivo della tutela perché tratta in termini più generali di persone fisiche, includendo non soltanto i singoli individui, ma anche i gruppi di persone. L'art. 50, par. 3, dell'AI Act, inoltre, obbliga chi implementa sistemi in grado di dedurre le emozioni all'informativa nei confronti degli interessati sul funzionamento del sistema stesso.⁵³

FRIA ha come obiettivo la tutela di tutti i diritti fondamentali, mentre la DPIA si interessa specificamente dei rischi legati alla tutela della privacy.

⁵⁰ Tra gli obblighi previsti per i sistemi di IA ad alto rischio si evidenziano la predisposizione di adeguati sistemi di valutazione e attenuazione dei rischi secondo un metodo «iterativo continuo» (art. 9 AI Act); essere caratterizzati da elevata qualità delle serie di dati che alimentano il sistema per ridurre al minimo i rischi di esiti discriminatori (art. 10 AI Act); registrazione delle attività per garantire la tracciabilità dei risultati (art. 12-19 AI Act, registrazione delle attività con logging continuo per tracciare il comportamento dell'algoritmo e auditabilità interna da parte delle autorità con possibilità di ricostruire e spiegare ogni decisione automatizzata e obbligo di informare l'utente che si sta interagendo con un sistema di IA); documentazione dettagliata che fornisca tutte le informazioni necessarie sul sistema e sulla sua finalità per consentire alle autorità di valutarne la conformità (art. 11 AI Act); informazioni chiare e adeguate fornite al deployer (art. 13 AI Act); adeguate misure di sorveglianza umana (art. 14 AI Act) e un elevato livello di robustezza, sicurezza informatica e accuratezza (art. 15 AI Act). Tali prescrizioni incombono sull'intera supply chain costituita da fornitori (produttori dei sistemi, anche situati al di fuori della UE), distributori di prodotti elettronici che importano (rimarchiandoli) dispositivi da Paesi UE o extra UE, distributori (installatori, system integrators e distributori di prodotti per la sicurezza che non rimarchiano i dispositivi) e utilizzatori (deployers, gli utenti finali, privati o pubblici che siano).

⁵¹ La definizione è peraltro coerente con i rilievi mossi nel parere congiunto EDPB-GEPD 5/2021 del 18 giugno 2021, sulla proposta di AI Act. Infatti, al punto 35 si legge che «(...) l'EDPB e il GEPD ritengono che l'utilizzo dell'IA per dedurre le emozioni di una persona fisica sia assolutamente inopportuno e dovrebbe essere vietato, ad eccezione di taluni casi d'uso ben specificati, ossia per finalità sanitarie o di ricerca (ad esempio pazienti per i quali il riconoscimento delle emozioni è rilevante), sempre applicando idonee tutele e, naturalmente, nel rispetto di tutte le altre condizioni e restrizioni relative alla protezione dei dati, compresa la limitazione delle finalità». Per ulteriori approfondimenti sul punto cfr., S. ORLANDO (a cura di), *Libertà e liceità del consenso nel trattamento dei dati personali*, Firenze, 2024.

⁵² Considerando 18 AI Act: «Il concetto si riferisce a emozioni o intenzioni quali felicità, tristezza, rabbia, sorpresa, disgusto, imbarazzo, eccitazione, vergogna, disprezzo, soddisfazione e divertimento. Non include stati fisici quali dolore o affaticamento, compresi, ad esempio, i sistemi utilizzati per rilevare lo stato di affaticamento dei piloti o dei conducenti professionisti allo scopo di prevenire incidenti. Non include nemmeno il semplice rilevamento di espressioni, gesti o movimenti facilmente riconoscibili, a meno che non siano utilizzati per identificare o dedurre emozioni. Tali espressioni possono essere espressioni facciali di base, come un cipiglio o un sorriso, o gesti come il movimento delle mani, delle braccia o della testa, o caratteristiche della voce di una persona, come un tono di voce alzato o un sussurro».

⁵³ Questo obbligo viene meglio definito nella sua portata attraverso lo studio del considerando 132 dell'AI Act che i loro dati biometrici sono oggetto di trattamento e che tali sistemi di IA possono «identificare o dedurre le emozioni o le intenzioni di tali persone».

In aggiunta l'art. 6 in combinato con l' Allegato III dell'AI Act inserisce i sistemi di riconoscimento delle emozioni impiegati nei contesti lavorativi, educativi, sanitari, nonché nei procedimenti di accesso ai servizi pubblici, tra i sistemi di IA ad alto rischio; e l'art. 5, par. 1, lett. f) dell'AI Act pone un divieto esplicito all'uso di sistemi di riconoscimento delle emozioni nelle istituzioni scolastiche e nei luoghi di lavoro, salvo eccezioni mediche o di sicurezza estremamente limitate, come i casi di utilizzo del sistema a scopo terapeutico⁵⁴. Evidentemente una disciplina di tal genere si pone in una posizione di rigidità rispetto all'impiego di tali sistemi. È proprio nelle more del Regolamento che si può leggere come la ragione di una tale scelta risiede nel fatto che tra le principali carenze di tali sistemi figurano l'affidabilità non dimostrata, la mancanza di specificità e la generalizzabilità limitata. Pertanto, i sistemi di IA che identificano o deducono le emozioni o le intenzioni delle persone fisiche sulla base dei loro dati biometrici possono portare a risultati discriminatori e possono ledere i diritti e le libertà delle persone interessate. Considerando lo squilibrio di potere nel contesto lavorativo o educativo, combinato con la natura invasiva di tali sistemi, questi ultimi potrebbero portare a un trattamento dannoso o sfavorevole di determinate persone fisiche o di interi gruppi.⁵⁵

4. L'integrazione "in concreto" della normativa vigente e le misure tecniche di protezione.

La compliance ai principi di legal design diventa cruciale in un contesto del genere, laddove il trattamento illecito dei dati può generare danni difficilmente riparabili, se solo si pensa ai riflessi che la divulgazione di informazioni sensibili può avere sulla vita personale, privata e pubblica di un individuo e dei suoi cari; sulle sue scelte e opportunità professionali; sulle sue abitudini di vita, ma anche al fatto che una rappresentazione probabilistica del sé emotivo di una persona ne può diventare la sua stessa dimensione esterna infedele.

Se la trasparenza è il principio generale cui è informato il GDPR⁵⁶, anche l'AI Act non è da meno e prevede che il funzionamento delle macchine sia improntato a tale principio, ma con la particolarità che non è richiesto che ci sia

⁵⁴ M. LEISER, *Psychological Patterns and Article 5 of the AI Act AI Powered Deceptive Design in the System Architecture and the User Interface*, in *Journal of AI Law and Regulation*, 2024, 1, 1, 5 - 23.

⁵⁵ In conformità al considerando 44 del Regolamento.

⁵⁶ M. E. KAMINSKI, *The Right to Explanation, Explained*, in *Berkeley Technology Law Journal*, Vol. 34, 1, 2019. L'A. evidenzia come ormai sia opportuno parlare ormai di "trasparenza qualificata". Questa si riscontra non soltanto negli obblighi informativi su logica e finalità del trattamento automatizzato ex art. 22 GDPR in capo al titolare, ma anche nel diritto di accesso ex art. 15 e nel diritto di opposizione di cui all'art. 21. Tali previsioni contribuiscono alla definizione del concetto di "accountability" in capo ai soggetti titolari del trattamento. In particolare, il titolare dovrà confermare all'interessato che sia o meno in corso un trattamento di dati personali e in caso affermativo informare l'interessato su finalità, categorie, destinatari, periodo di conservazione, diritti di rettifica e/o cancellazione, origine dei dati (se non raccolti in modo diretto) e logica dei processi decisionali automatizzati (profilazione), con diritto a una copia gratuita dei dati trattati; presa consapevolezza del trattamento in corso l'interessato potrà opporvisi ex art. 21; ai sensi dell'art. 22 invece il titolare dovrà garantire all'interessato la possibilità di ottenere l'intervento umano da parte del titolare del trattamento, di esprimere la propria opinione e di contestare la decisione, nei casi in cui l'interessato abbia prestato consenso esplicito al trattamento automatizzato o il trattamento automatizzato sia necessario per la conclusione o esecuzione di un contratto tra titolare e interessato. Cfr., tra gli altri, R. Montinaro, *op. cit.*, ibidem.

in corso un trattamento di dati personali affinché si applichino le misure. Occorre infatti garantire trasparenza tecnica (accesso a documentazione tecnica, logiche algoritmiche, pesi e dataset) e comprensibilità e interpretabilità per l'utente (spiegazioni accessibili, chiare e contestualizzate per persone non esperte) come preconditione per l'operatività dei sistemi di IA⁵⁷.

Un sistema che non consente trasparenza e spiegazioni verificabili e interpretabili non può essere conforme all'AI Act⁵⁸. Sarebbe però riduttivo declinare il principio di trasparenza e spiegabilità in una misura esclusivamente tecnica.

La prima sfida alla trasparenza deriva infatti dalla formulazione giuridica aperta dei requisiti tecnici, che nel diritto dell'UE sono tipicamente descritti in termini astratti. Per rendere il funzionamento del sistema ad alto rischio «sufficientemente trasparente da consentire ai deployer di interpretare l'output del sistema e utilizzarlo adeguatamente»⁵⁹, un fornitore può costruire il proprio sistema in maniera di partenza interpretabile oppure può anche soddisfare tale requisito utilizzando un modello che consenta agli utenti di comprendere in che modo i parametri di input sono collegati all'output o fornendo linee guida contestuali sull'interpretazione dell'output stesso. Tale vaghezza nella formulazione normativa è il prezzo che il legislatore europeo paga al suo approccio tecnologicamente neutrale⁶⁰ e ciò lascia ampia libertà di azione agli organismi di regolamentazione e certificazione e agli attori pubblici e privati della filiera.

Tale modello di regolazione “aperto” si completa con la soft law⁶¹, al fine di meglio specificare i contenuti prescrittivi in concreto e, in un'ottica di “law in

⁵⁷ Pur non dando delle definizioni esatte delle terminologie utilizzate, quali in particolare la trasparenza, la spiegabilità e l'interpretabilità, queste emergono dalle prescrizioni contenute nel corpo normativo. In tal senso si vedano, anche ma non solo (in quanto gli obblighi di trasparenza riguardano pressoché tutti i sistemi di IA nella piramide del rischio, ad eccezione dei sistemi a rischio nullo) gli artt. 13 e 16 dell'AI Act indirizzati ai fornitori di sistemi ad alto rischio, così come gli obblighi di conservazione di cui agli artt. 18 e 19.

⁵⁸ Tali obblighi si completano con quelli previsti dal GDPR già accennati *supra*: informativa obbligatoria sull'uso dell'IA (artt. 13-14), in quanto l'interessato deve sapere che una decisione è stata presa in una modalità automatizzata e con quali logiche; diritto a non essere sottoposto a decisioni automatizzate se non c'è supervisione umana o base giuridica chiara e diritto a ricevere spiegazioni ai sensi degli artt. 15 e 22, in quanto la persona deve essere messa nelle condizioni di comprendere perché ha ricevuto un certo output da parte della macchina.

⁵⁹ Art. 13 dell'AI Act.

⁶⁰ È una costante dell'approccio regolatorio europeo che si conferma anche nelle più recenti e note regolamentazioni come DSA e DMA, e nell'AI Act, sebbene non ve ne sia all'interno una esplicita menzione. Si preferisce un approccio regolatorio “ad ampio raggio” per principi e regole, che lasci uno spazio successivo per la normazione di dettaglio, basata sull'intervento di organismi di certificazione e autorità indipendenti che si occupano della normazione tecnica. In tal modo il legislatore europeo ritiene di poter adottare uno strumento più aderente all'evoluzione tecnologica ed al contempo efficiente (grazie al fatto che gli strumenti di soft law possono essere modificati e aggiornati attraverso procedimenti più semplici rispetto all'hard law); in poche parole l'UE ritiene che tale approccio sia più efficiente. Cfr., M. ALMADA, *Two dogmas of technology-neutral regulation*, 2024, consultabile su SSRN, <http://dx.doi.org/10.2139/ssrn.4953377>, M. D. BIRNHACK, *Reverse Engineering Informational Privacy Law*, in *Yale Journal of Law & Technology*, 2012, XV, 1, 24–91, C. ROSCA, *Digital Arms for Digital Consumer Harms: Mapping Legal and Technical Solutions for Dark Patterns*, 2024, EU Consumer Law [PhD in Law], Maastricht University.

⁶¹ Nel luglio 2025, la Commissione ha introdotto 3 strumenti chiave per supportare lo sviluppo e l'impiego responsabile dei modelli GPAI: le Linee Guida per gli obblighi per i fornitori di modelli GPAI; il Codice di Pratica GPAI che offre indicazioni pratiche per aiutare i fornitori a rispettare gli obblighi previsti dall'AI Act in materia di trasparenza, diritto d'autore, sicurezza e protezione e il modello per il riepilogo pubblico del contenuto dei dati di addestramento dei modelli GPAI che richiede ai fornitori di fornire una panoramica dei

practice”, tribunali e autorità amministrative possono cercare di specificare ulteriormente i requisiti imposti dalla legge attraverso richieste di disclosure (in caso di conformità agli standard tecnici armonizzati dei sistemi “high risk” si dovrà procedere con un’ulteriore inversione dell’onere della prova in capo alle autorità). Sono però molteplici le casistiche in cui tale richiesta può essere legittimamente disattesa da autorità di regolamentazione e attori regolamentati: trade secrets, tutela del segreto commerciale dei fornitori di IA; segreto di Stato; ulteriori ostacoli alla divulgazione possono essere creati attraverso mezzi contrattuali, ad esempio il fornitore di un sistema di IA può imporre a chi implementa tale sistema clausole che impediscono la revisione delle informazioni relative alla sua logica decisionale e/o ai suoi parametri; infine, la divulgazione può essere limitata da disposizioni di legge che tutelano terzi, come nei casi in cui i modelli che vengono addestrati su dataset di grandi dimensioni che possono includere dati personali di persone fisiche che non hanno dato consenso al trattamento.

Così è chiaro come il termine black box, generalmente collegato all’opacità algoritmica, non si esaurisce in quest’ultima, ma trova ulteriore espressione nell’interazione fra diritto alla trasparenza e altri diritti, che posti in bilanciamento possono prevalere⁶².

Per rendere maggiormente effettivo il principio di trasparenza occorrerebbe che i sistemi di chatbot siano pensati fin dal principio in conformità alle regole del GDPR e dell’AI Act. In tal senso, assumono cruciale rilevanza le cd. misure tecniche di protezione: soluzioni tecnologiche e strutturali che riducono o eliminano i rischi di danno realizzando in concreto il principio di trasparenza⁶³.

È fondamentale che, ad esempio, i dati trattati vengano conservati esclusivamente per il tempo necessario all’esecuzione della chat, e che non vengano trattenuti oltre. Un’ulteriore misura riguarda la raccolta dei dati, che può essere effettuata in forma anonimizzata, in quanto la pseudonimizzazione non garantisce, soprattutto in caso di grandi set di dati, che non ci sia la possibilità di risalire al titolare dei dati.

Molte aziende oggi si affidano a strumenti di post-hoc explainability⁶⁴ per rendere interpretabili anche i modelli black box. Un’altra strada è quella dell’explainability by design: modelli intrinsecamente interpretabili,

dati utilizzati per addestrare i propri modelli, comprese le fonti da cui sono stati ottenuti i dati (tra cui i grandi dataset e i principali nomi di dominio).

⁶² Del resto, è osservazione condivisa, e già evidenziata dagli Illustri Maestri Bianca e Busnelli, che il diritto alla privacy sia espressione di un modello “relazionale”, in particolare in bilanciamento con il diritto all’informazione. Cfr. C. M. BIANCA, F. D. BUSNELLI, *op. cit.*, 31-35 e 14.

⁶³ In tale prospettiva è stato sviluppato il progetto M.E.I. nel corso della VI edizione della FIVEG Academy dell’Università Federico II di Napoli, di cui mi sia consentito di ringraziare il gruppo di lavoro e in particolare i docenti A. M. Tulino e A. Origlia per l’occasione di confronto concessami. Per ulteriori approfondimenti cfr., <https://www.5gacademy.unina.it/chi-siamo/contatti.html>.

⁶⁴ LIME e SHAP sono metodi per interpretare modelli di apprendimento automatico, in particolare le cd. “black box”. LIME (Local Interpretable Model-agnostic Explanations) crea spiegazioni locali attraverso un modello semplice e interpretabile intorno a una singola previsione. SHAP (SHapley Additive exPlanations) utilizza i valori di Shapley per fornire una spiegazione più completa e coerente che si può anche estendere a una visione globale del modello, cfr. <https://medium.com/@pacosun/trusting-shap-and-lime-the-right-way-ccab78f0316b> (accesso effettuato in data 8.11.2025).

caratterizzati magari da una minore accuratezza, ma da una maggiore interpretabilità⁶⁵.

È inoltre possibile differenziare su più livelli l'apprendimento della macchina, sulla base del modello di *federated learning*: l'addestramento della macchina è centralizzato, ma basato su dati trattati in forma decentralizzata in cui gli attori della filiera non condividono dati riconoscibili a posteriori, ma trattati solo in forma anonima e aggregata, sul modello della GBoard di Google⁶⁶.

Ai fini dell'accesso ai dati sono poi attuabili diverse misure di protezione aggiuntiva, come la crittazione dei dati stessi o l'utilizzo di VPN protette per accedere al cloud. Tali misure risultano fondamentali in una prospettiva di mitigazione del rischio, ma ciò che in questo elaborato si vuole esaminare è se, al netto di tali misure, i benefici che tali sistemi apportano alla collettività possano superare le esternalità negative.

5. Conclusioni e prospettive evolutive.

Passiamo in rassegna l'itinerario argomentativo sin qui evidenziato per avviarci verso la conclusione dell'elaborato. Abbiamo visto come un chatbot possa essere qualificato come sistema di riconoscimento delle emozioni quando inferisce una o più delle emozioni o intenzioni umane mediante l'analisi dei dati biometrici o personali di una o più persone fisiche.

Questo implica che la macchina presenta un trattamento tecnico specifico volto a ricavare caratteristiche fisiche, fisiologiche o comportamentali che consentano o confermino l'identificazione univoca di quella persona o di quel gruppo di persone⁶⁷ e che il trattamento deve essere specificamente orientato a tale scopo⁶⁹. Non, perciò, le emozioni in quanto tali, ma le categorie particolari di dati che possano emergere dalla profilazione delle emozioni dovranno essere tutelate in quanto sensibili.

Cioè, è necessario distinguere tra il vero e proprio dato biometrico (il risultato del trattamento, come ad esempio l'impronta del volto o quella digitale) e la fonte dello stesso (la fotografia, il dito, il timbro vocale) che può anche essere oggetto di rilevamento automatizzato delle caratteristiche fisiche di un individuo. Solo se la finalità è quella di identificare univocamente una persona fisica e il trattamento tecnico cui sono sottoposte le informazioni

⁶⁵ In termini tecnici si parla di trade off tra accuratezza e interpretabilità date una funzione obiettivo di una certa complessità e una quantità adeguata di dati disponibili. Cfr., S. RUSSELL, P. NORVIG, *Artificial Intelligence: A Modern Approach*, Global Edition 4, Pearson, 2021.

⁶⁶ Un chatbot viene individuato e realizzato secondo dei dettami tecnici specifici che consentono un'archiviazione dei dati differenziata su due scale: una in locale e una decentralizzata. Il server centrale raccoglie lo storico delle conversazioni schematizzate secondo informazioni rilevanti per l'addestramento in forma totalmente anonimizzata; in delocalizzato vengono invece raccolti i dati delle anomalie riferite alla specifica macchina. Un ulteriore elemento fondamentale per il rispetto della reale volontà dell'utente nell'interazione con la macchina è che, a fronte delle domande poste nella forma più naturale possibile dalla macchina stessa, questa accetti che a determinate domande l'utente non intenda rispondere, e nonostante ciò riesca a elaborare una risposta credibile ai suoi quesiti. Si tratta, in gergo tecnico, del funzionamento del chatbot anche in presenza dei cd. "missing values".

⁶⁷ Art. 4, punto 14, GDPR.

⁶⁹ Il gruppo di lavoro art. 29 nel già citato parere n. 3/2012 esprimeva preoccupazione in quanto alcuni sistemi possono raccogliere segretamente informazioni relative agli stati emotivi o alle caratteristiche fisiche e rivelare informazioni sulla salute, determinando un trattamento dei dati non proporzionato nonché il trattamento di dati sensibili.

è idoneo a estrarre i dati biometrici allora si potrà concludere che siamo in presenza di dati biometrici soggetti alla disciplina di cui all'art. 9 GDPR: l'analisi in tal senso potrà essere effettuata solamente in concreto, caso per caso⁷⁰. Conformemente si è espressa l'Autorità Garante per la privacy italiana⁷¹, stabilendo che i dati acquisiti per il riconoscimento via webcam non sono dati biometrici e quindi non richiedono particolari cautele. Perché si possa parlare di trattamento di dati biometrici occorre, invece, che la verifica dell'identità sia automatizzata, tramite l'ausilio di appositi strumenti software o hardware.

Ad una posizione parzialmente difforme sono pervenuti gli Ermellini sostenendo che «ricorre un trattamento di dati biometrici quando gli stessi sono ottenuti mediante un trattamento tecnico automatizzato specifico, realizzato con un software che, sulla base di riprese e analisi delle caratteristiche fisiche, fisiologiche o comportamentali di una persona fisica, le elabora, evidenziando comportamenti o elementi anomali, e che perviene a un esito conclusivo, costituito da un elaborato video/foto che consente (o che conferma) l'identificazione univoca della persona fisica, restando irrilevante la circostanza che l'esito finale del trattamento sia successivamente sottoposto alla verifica finale di una persona fisica»⁷². Nel caso specifico la Corte precisa che, se le riprese video di un esame non hanno la sola funzione di documentare la prova di esame, ma si connotano per la contestuale elaborazione e selezione del materiale raccolto, che converge nella individuazione ed alla segnalazione di comportamenti anomali, attraverso la produzione del video finale, siamo in presenza di un trattamento di dati biometrici.

A *contrariis*, un sistema non è considerato sistema di riconoscimento delle emozioni, ai sensi della disciplina introdotta dall'AI Act e completata con le disposizioni date dal GDPR, nel momento in cui: rilevi determinati movimenti o espressioni facciali, senza inferire una precisa emozione e senza processare dati biometrici; identifichi stati fisici non direttamente connessi ad un'emozione, quali dolore o affaticamento (è il caso di scuola dei sistemi di riconoscimento della stanchezza alla guida); per quanto riguarda l'eventuale captazione audio, si limita a rilevare determinate caratteristiche della voce senza analizzarla in modo tale da poterla ricondurre a un singolo individuo (quindi viene garantita anonimizzazione). In questo spazio operativo va individuata la possibilità di intervenire attraverso l'implementazione di sistemi di apprendimento e rilevazione avanzati.

Concludendo, attraverso questo elaborato si è cercato di dare una prima panoramica e una iniziale spiegazione del perché l'utilizzo di un chatbot in grado di dedurre le emozioni è circoscritto a casi eccezionali e, considerando il combinato disposto con l'art. 9 GDPR, al caso in cui non ci siano opzioni meno invasive per raggiungere la finalità del trattamento (e senza tralasciare il prerequisite dell'implementazione delle misure tecniche di protezione più adeguate).

⁷⁰ P. QUINN, G. MALGIERI, *The Difficulty of Defining Sensitive Data – The Concept of Sensitive Data in the EU Data Protection Framework*, in *German Law Journal*, 2021, XXII, 1583-1584.

⁷¹ Verifica preliminare. Riconoscimento via webcam dei partecipanti a corsi di formazione in diretta streaming - 26 luglio 2017, Registro dei provvedimenti n. 345 del 26 luglio 2017. Consultabile su <https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/6826368>.

⁷² Cass. Civ., sez. I, ord. 13.05.2024, n. 12967, in *Italggiure Web*.

I sistemi che inferiscono le emozioni lo fanno instaurando un processo ontologicamente deduttivo, in cui a partire da alcuni tratti individuati e confrontati con grandi masse di dati, viene tracciato un profilo dell'utente. Tale logica è per certi versi distante dal garantire che gli elementi profilati rappresentino fedelmente le convinzioni profonde e le emozioni provate nell'intimo dalla persona, giustificandosi così un approccio particolarmente cauto del legislatore comunitario nello sdoganamento di tali sistemi ad ampio spettro e nei contesti più sensibili.⁷³

Dal momento che non siamo più dinanzi a un mero trattamento di informazioni, ma a un'interferenza profonda con la dimensione intima della persona, la sua sfera emotiva, il suo diritto a essere sé stessa senza essere analizzata, profilata o manipolata sulla base di stati psicologici transitori⁷⁴, la sfida per il giurista non consiste soltanto nell'applicare regole già esistenti, ma nel contribuire a adattarle e, se necessario, innovarle, affinché il progresso tecnologico resti conforme a quei principi di dignità, libertà e autodeterminazione che sono al centro del diritto europeo e internazionale.

La posta in gioco non è meramente tecnica o giuridica, ma profondamente politica e culturale: si tratta di decidere quale equilibrio intendiamo stabilire tra innovazione e tutela dei diritti fondamentali, mantenendo centrale il prisma della tutela dell'autonomia e della dignità delle persone⁷⁵.

Se sapremo costruire ambienti salubri, capaci di garantire un miglioramento della qualità della vita delle persone attraverso l'implementazione di tecnologie affidabili e attendibili, potremo far sì che l'intelligenza artificiale diventi una leva di progresso, senza sacrificare i valori che fondano i nostri ordinamenti.⁷⁶

⁷³ Ciò non esclude, come si può leggere nella stessa normativa comunitaria, che a eccezionali condizioni questi sistemi possano risultare un elemento di supporto all'uomo, come già evidenziato nel paragrafo introduttivo, quando si è fatto riferimento ai casi di macchine per l'apprendimento che possono essere utilizzate nei contesti educativi in cui sono presenti particolari forme patologiche, come nelle situazioni di studenti non verbali.

⁷⁴ Occorre soffermarsi, in particolare, sulla situazione dei soggetti più vulnerabili, come i minori, i quali possono instaurare con le macchine dotate di IA relazioni di tipo empatico, quasi amicale, tali da incidere in maniera significativa sulla loro capacità decisionale. Cfr. M. C. GAETA, *Smart toys and minors' protection in the context of the Internet of everything*, in *EJPLT*, 2020, II.

⁷⁵ In tal senso, è illuminante un passaggio del già citato Commentario, in cui si evidenzia come la tutela della riservatezza, così come quella dell'identità personale e del diritto alla protezione dei dati personali sono filtrate attraverso il principio della «"dignità dell'interessato" [...]»; principio fondamentale che consente un'opera di disarticolazione della generale categoria dei dati personali in funzione della individuazione di tutele graduate a seconda dei valori più o meno forti che un generico riferimento alla privacy necessariamente sottende». Cfr. C. M. BIANCA, F. D. BUSNELLI, *op. cit.*, 58 - 59

⁷⁶ «[...] the design, development, and deployment of AI systems in ways that prevent, mitigate or resolve problems adversely affecting human life and/or the wellbeing of the natural world, and/or enable socially preferable and/or environmentally sustainable developments», cfr., L. FLORIDI, et al., *How to Design AI for Social Good: Seven Essential Factors*, in *Sci Eng Ethics*, 2020, vol. 26, III, 1771-1796. Consultabile su <https://doi.org/10.1007/s11948-020-00213-5>.