L'IRCOCERVO
PRIMA RIVISTA ELETTRONICA ITALIANA DI METODOLOGIA GIURIDICA - TEORIA GENERALE DEL DIRITTO E DOTTRINA DELLO STATO
FIRST ITALIAN DIGITAL JOURNAL OF LEGAL METHODOLOGY - GENERAL THEORY OF LAW AND DOCTRINE OF THE STATE

Consenso e capacità giuridica nell'era dell'IA neuroassistita: profili bioetici e implicazioni dell'art. 14 *AI Act*

Fabrizio Corona e Alessia Del Pizzo*

Università Giustino Fortunato
Università LUISS Guido Carli

Abstract: **Consent and Legal Capacity in the Era of Neuro-assisted AI: Bioethical Profiles and Implications of Article 14 of the AI Act**
This paper examines the legal and ethical challenges posed by the use of brain-computer interfaces (BCIs) in the decision-making processes of individuals with severe communicative disabilities, particularly in cases of locked-in syndrome (LIS) and completely locked-in state (CLIS). These technologies, by enabling the expression of will through neural signals, call for a critical reassessment of the traditional foundations of will manifestation and deeply question the very notion of legal capacity. The issue is not merely whether individuals can perform dispositive acts, but whether legal subjectivity can be upheld without reliance on conventional expressive channels. The paper explores the possibility of assigning legal value to expressions of will entirely mediated by algorithms, highlighting the challenges related to computational opacity and neurophysiological variability. In this context, Article 14 of Regulation (EU) 2024/1689 on Artificial Intelligence (AI Act) is analyzed. This provision introduces the principle of meaningful human oversight for high-risk systems such as BCIs. Such oversight is interpreted not as a mere technical requirement, but as an essential ethical and legal safeguard to ensure authenticity, responsibility, and self-determination.

Keywords: Brain-Computer Interface, Algorithmic Legality, Human Supervision, AI, Machine Learning.

Sommario: 1. Le *BCI* nel contesto clinico: tecnologie abilitanti della volontà giuridica – 2. Dare voce alla volontà: il riconoscimento giuridico dell'intenzionalità espressa tramite *BCI* – 2.1. *CLIS*, *BCI* e legalità algoritmica: dalla volontà immota alla trasparenza computazionale – 3. Il concetto di supervisione umana ai sensi dell'art. 14 – 3.1. La supervisione umana nella *governance* dei sistemi di IA basati su tecnologie *BCI* – 3.2. Il principio della supervisione umana nell'*AI Act* – 4. I limiti della supervisione umana e le prospettive regolatorie – 5. Riflessioni Conclusive.

* § 1, 2, 2.1. sono redatti da Alessia Del Pizzo. § 3, 3.1., 3.2., 4 sono redatti da Fabrizio Corona. Il § 5 è frutto delle considerazioni comuni.

© L'Ircocervo, 24 (2025) n. 2 ISSN 1722-392X
Special issue: A.C. Amato Mangiameli, P. Becchi, P.B. Helzel, G. Tarantino (eds.),
Bioethics and Biolaw Facing the Current Challenges,
https://doi.org/10.57581/LIRCOCERVO.24.2025.2.06

1. Le *BCI* nel contesto clinico: tecnologie abilitanti della volontà giuridica

Negli ultimi due decenni, l'evoluzione delle interfacce cervello-computer (*Brain-Computer Interfaces*, *BCI*) ha significativamente trasformato il panorama clinico, offrendo possibilità inedite a pazienti in stato di grave disabilità motoria o comunicativa.

Queste tecnologie, concepite per intercettare segnali elettrofisiologici emessi dal cervello e tradurli in comandi digitali interpretabili da dispositivi esterni – come *computer*, sintetizzatori vocali, ausili motori o ambienti domotici – si fondano sul presupposto che l'intenzionalità cognitiva sia accessibile, rilevabile e formalizzabile anche in assenza di espressione corporea[1].

L'interesse crescente verso queste soluzioni è stato sostenuto da una convergenza di fattori: una maggiore comprensione delle dinamiche neurali, la disponibilità di tecnologie informatiche sempre più potenti e a basso costo, e un'attenzione più consapevole verso le esigenze delle persone con disabilità gravi.

In particolare, i programmi di ricerca si sono concentrati sullo sviluppo di sistemi di comunicazione e controllo destinati a soggetti affetti da sindrome *locked-in* (*LIS*), una condizione caratterizzata da paralisi motoria completa e, in molti casi, da insufficienza respiratoria compensata tramite ventilazione artificiale. Nei casi più gravi, si parla di sindrome completamente *locked-in* (*CLIS*), in cui si verifica anche la paralisi totale dei muscoli oculari, rendendo impossibile qualsiasi forma di comunicazione convenzionale[2].

In questo contesto, ciò che tradizionalmente veniva inteso come "comunicazione[3]" – ovvero un atto espressivo mediato da linguaggio verbale, gesti,

[1] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, T.M. Vaughan, "Brain–computer interfaces for communication and control", in *Clinical Neurophysiology*, 113 (2002), n. 6, pp. 767-791, doi: 10.1016/s1388-2457(02)00057-3.

[2] La *LIS* descrive una condizione in cui la coscienza è pienamente integra, ma il paziente è totalmente paralizzato, afonico e incapace di comunicare verbalmente, potendo talvolta utilizzare solo i movimenti oculari. Le cause più comuni includono malattie del motoneurone, come la SLA, lesioni pontine da ictus, traumi cranici, anossie cerebrali e patologie cerebrovascolari, tutte responsabili dell'interruzione delle vie motorie pur lasciando intatta la funzione corticale. E. Casanova, R.E. Lazzari, S. Lotta, A. Mazzucchi, "Locked-in syndrome: improvement in the prognosis after an early intensive multidisciplinary rehabilitation", in *Archives of Physical Medicine and Rehabilitation*, 84 (2003), pp. 862-867.

[3] La comunicazione è una delle facoltà fondative della condizione umana, intesa non solo come scambio di informazioni, ma come processo simbolico e relazionale che costruisce significati condivisi. Derivante dal latino *communicatio*, ovvero "mettere in comune", essa implica almeno due soggetti, un messaggio e un codice comprensibile da entrambe le parti. Lungi dall'essere un atto meramente tecnico, la comunicazione è intrinsecamente situata: è influenzata dai contesti storici, sociali, culturali e affettivi nei quali si realizza. Nel paradigma contemporaneo, la comunicazione è divenuta sempre più ibrida, articolata tra linguaggi verbali e non verbali, media fisici e digitali, interazioni sincrone e asincrone. Tuttavia, essa non è mai un processo unilaterale, bensì un'attività reciproca, in cui emittente e ricevente sono co-autori del significato: in gioco non c'è soltanto la produzione/trasmissione di un messaggio, bensì la costruzione del suo significato. N. Martini, V. Mastronardi, "Il mondo della comunicazione: comunicazione non verbale (CNV), programmazione

movimenti o segnali corporei osservabili – può oggi essere sostituito dalla sola attività cerebrale, rilevata attraverso sensori neurofisiologici, codificata in segnali digitali e infine convertita in output intelligibile da parte di un sistema esterno. Questo processo avviene grazie all'impiego di algoritmi di intelligenza artificiale, i quali svolgono un ruolo decisivo nell'analisi, classificazione e interpretazione dei neurodati[4]. Questi algoritmi, addestrati tramite tecniche di *machine learning* e *deep learning*, sono in grado di individuare correlazioni tra specifici stati mentali e *output* comportamentali – ad esempio, l'intenzione di muovere un arto o di selezionare una lettera su uno schermo – e di trasformare tali correlati in istruzioni operative per dispositivi esterni[5]. La capacità di questi sistemi di apprendere dalle esperienze neurali individuali consente una personalizzazione dinamica, rendendo le *BCI* sempre più precise, adattive e in grado di riconoscere anche segnali deboli o ambigui.

Si configura così una nuova forma di comunicazione, non più vincolata alle modalità espressive tradizionali del corpo, ma resa possibile da un'interazione diretta tra mente e macchina: è il cervello che "parla" attraverso segnali neurali, mentre l'algoritmo "ascolta", interpreta e traduce tale attività mentale in comandi operativi[6]. Questa trasformazione rappresenta una discontinuità radicale, tanto sul piano epistemologico quanto su quello antropologico. La comunicazione non necessita più del corpo come tramite necessario: può avvenire anche in condizioni di immobilità assoluta, mediante un'interfaccia che connette direttamente il pensiero alla tecnologia. In questa prospettiva, il paradigma classico della *res extensa* cartesiana – secondo cui il corpo costituisce l'unico tramite attraverso cui la mente può agire nel mondo – viene profondamente riconsiderato. La *res cogitans*, tradizionalmente relegata alla sfera dell'interiorità, acquisisce oggi una dimensione esterna, operativa e performativa[7]. Si rovescia così la gerarchia cartesiana: non è più il corpo a rendere possibile l'azione, ma la mente, grazie alla mediazione tecnologica, può esprimersi e incidere direttamente sulla realtà. Questo mutamento apre a una nuova concezione della soggettività giuridica, intesa come espressione immediata della volontà mentale, capace di produrre effetti giuridici anche in

neuro-linguistica (PNL) e strategie di persuasione", in *Rivista di Psicopatologia Forense, Medicina Legale, Criminologia*, 25 (2020), p. 547.

[4] I neurodati rappresentano la materia prima dell'attività cerebrale: si tratta di segnali elettrici, magnetici o metabolici che, sebbene inizialmente privi di una struttura informativa compiuta, possono essere trasformati, attraverso avanzati processi di elaborazione algoritmica e tecniche di *machine learning*, in vere e proprie informazioni mentali. M. Ienca, *Intelligenza². Per un'unione di intelligenza naturale e artificiale*, Rosenberg & Sellier, Torino, 2019.

[5] H. Zhang, L. Jiao, S. Yang, H. Li, X. Jiang, J. Feng, S. Zou, Q. Xu, J. Gu, X. Wang, B. Wei, "Brain-computer interfaces: the innovative key to unlocking neurological conditions", in *International Journal of Surgery*, 110 (2024), n. 9, pp. 5745-5762.

[6] A. Del Pizzo, "Beyond the Cure: Neurotechnology and Law in the Age of Cognitive Unravelling", in *Ragion Pratica. Rivista semestrale*, (2025), n. 1, pp. 73-91.

[7] Cfr. R. Descartes, *Discorso sul metodo*, trad. it. di E. Garin, Laterza, Roma-Bari, 1994 (ed. orig. 1637); F. Brentano, *Psychologie vom empirischen Standpunkte*, Duncker & Humblot, Lipsia, 1874.

assenza di corporeità attiva.

Il caso dei pazienti in stato di *completely locked-in* (*CLIS*), condizione in cui ogni funzione motoria è completamente compromessa ma la coscienza e le funzioni cognitive restano intatte, costituisce il banco di prova più estremo queste tecnologie. In questi soggetti, ogni modalità comunicativa tradizionale – verbale, gestuale, anche oculare – risulta inaccessibile, relegando la soggettività cosciente a una condizione di isolamento assoluto. Si potrebbe presumere che una simile condizione, caratterizzata dall'intrappolamento nel proprio corpo, sia inevitabilmente associata a una scarsa qualità della vita e a un desiderio diffuso di morte medicalmente assistita. Eppure, sorprendentemente, numerosi studi hanno dimostrato che molti pazienti *CLIS*, pur in questa condizione estrema, esprimono un forte desiderio di vivere e attribuiscono significato alla propria esistenza, a patto di poter stabilire una forma di comunicazione con il mondo esterno[8].

In questo contesto, la *BCI* non si limita a essere un ausilio tecnico, ma si configura come l'unico dispositivo capace di riattivare un legame intenzionale tra l'individuo e il mondo: un mezzo attraverso il quale è possibile rispondere a una domanda, esprimere un consenso, comunicare una preferenza o rifiutare un trattamento sanitario[9].

Il valore di questa tecnologia, dunque, trascende la dimensione assistiva: essa si trasforma in una protesi cognitiva e giuridica, in grado di restituire al soggetto non solo un canale di comunicazione, ma anche la possibilità di esercitare i propri diritti fondamentali. L'attività cerebrale, tradotta in segnali elettrici e poi in comandi comprensibili, diventa una nuova forma di linguaggio operativo. In altre parole, il pensiero – tradizionalmente confinato alla dimensione interiore – acquisisce una forza performativa, capace di generare effetti giuridici nel mondo esterno. Il significato antropologico e normativo di questa trasformazione è profondo: si passa dalla mera sopravvivenza clinica alla riabilitazione della capacità d'agire, intesa come espressione della libertà individuale.

Da un punto di vista neurotecnico, le B*CI* impiegate nei pazienti in stato di *CLIS* si basano su diverse strategie di rilevamento e classificazione dei segnali neurali. Le metodologie più consolidate prevedono l'utilizzo di segnali elettroencefalografici (EEG), in particolare quelli generati dai potenziali evocati P300, che corrispondono a risposte neurali a stimoli esterni rilevanti. Tali risposte

[8] Si vedano: D. Lulé, C. Zickler, S. Häcker, M.-A. Bruno, A. Demertzi, F. Pellas, S. Laureys, A. Kübler, "Life can be worth living in locked-in syndrome", in *Progress in Brain Research*, 117 (2009), pp. 339-351; M.-A. Bruno, J.-L. Bernheim, D. Ledoux, F. Pellas, A. Demertzi, S. Laureys, "A survey on self-assessed well-being in a cohort of chronic locked-in syndrome patients: happy majority, miserable minority", in *BMJ Open*, 1 (2011); M.-C. Rousseau, K. Baumstarck, M. Alessandrini, V. Blandin, T. Billette de Villemeur, P. Auquier, "Quality of life in patients with locked-in syndrome: evolution over a 6-year period", in *Orphanet Journal of Rare Diseases*, 10 (2015).

[9] A. Kübler, N. Birbaumer, "Brain-computer interfaces and communication in paralysis: Extinction of goal directed thinking in completely paralysed patients?", in *Clinical Neurophysiology*, 119 (2008), n. 11, pp. 2658-2666.

vengono generalmente elicitate attraverso l'impiego del cosiddetto "*oddball paradigm*"[10], un paradigma sperimentale in cui viene presentata una sequenza di stimoli standard interrotta da stimoli rari (*target*) che, se riconosciuti dal soggetto, generano una deflessione positiva nel tracciato EEG a circa 300 millisecondi.

Questo schema è alla base di molte interfacce *BCI* orientate alla comunicazione, come nel caso del noto P300 *speller*, in cui la selezione mentale di una lettera o simbolo attiva la risposta cerebrale rilevabile dal sistema[11].

Oltre al P300, si utilizzano anche le bande μ e β, associate rispettivamente all'immaginazione motoria e all'attività cognitiva. Accanto a queste tecniche, si è affermato l'uso della spettroscopia nel vicino infrarosso (*fNIRS*), che consente di rilevare variazioni emodinamiche corticali legate a stati di attenzione e decisione volontaria, anche in soggetti completamente paralizzati. Tale approccio si è rivelato particolarmente utile nei pazienti in stato di *completely locked-in*, nei quali la qualità dei segnali EEG può risultare insufficiente o compromessa[12].

Un ulteriore contributo in questa direzione è rappresentato dallo sviluppo di modelli ibridi *EEG-fNIRS*, che combinano l'alta risoluzione temporale dell'*EEG* con la sensibilità emodinamica della *fNIRS*. Tali soluzioni si sono dimostrate promettenti nel migliorare la precisione dell'interpretazione neurale e nel ridurre significativamente la percentuale di falsi positivi – una criticità centrale, soprattutto nei contesti in cui il riconoscimento dell'intenzione dell'utente comporta conseguenze giuridiche rilevanti, come l'espressione di un consenso informato o di volontà testamentarie[13].

Un esempio significativo della diversificazione dei canali sensoriali impiegati nei sistemi BCI è offerto da uno studio del 2014 che ha dimostrato per la prima volta la fattibilità di un'interfaccia cervello-computer basata su stimolazione vibrotattile, consentendo una forma basilare di comunicazione anche in pazienti con *deficit* visivi o uditivi. In questo esperimento, sei pazienti in stato di *LIS* sono stati sottoposti a un paradigma *oddball* somatosensoriale, in cui stimolatori vibranti erano applicati ai polsi: i partecipanti dovevano focalizzarsi mentalmente sul polso designato (ad esempio, destro per "sì", sinistro per "no") e contare le vibrazioni target. L'onda P300 è stata rilevata in tutti i pazienti con tre stimolatori e in cinque su sei con due stimolatori. Nella fase comunicativa, un partecipante ha raggiunto il 100% di accuratezza nelle risposte binarie e altri hanno superato significativamente il livello del caso, aprendo così alla possibilità concreta di una comunicazione

[10] L.A. Farwell, E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials", in *Electroencephalography and Clinical Neurophysiology*, 70 (1988), n. 6, pp. 510-523.
[11] *Ibidem.*
[12] S. Rezvani, S. Hosseini-Zahraei, A. Tootchi, *et al.*, "A review on the performance of brain-computer interface systems used for patients with locked-in and completely locked-in syndrome", in *Cognitive Neurodynamics*, 118 (2024), pp. 1419-1443.
[13] C. Guger, R. Spataro, B.Z. Allison, A. Heilinger, R. Ortner, W. Cho, *et al.*, "Complete locked-in and locked-in patients: Command following assessment and communication with vibro-tactile P300 and motor imagery brain-computer interface tools", in *Frontiers in Neuroscience*, 11 (2017).

volontaria anche in assenza di canali sensoriali tradizionali[14].

La letteratura neuroscientifica degli ultimi anni ha documentato casi in cui pazienti *CLIS*, nonostante l'assenza di qualsiasi residuo motorio, sono riusciti a rispondere correttamente a domande personali o clinicamente rilevanti utilizzando esclusivamente l'attività cerebrale. In particolare, i paradigmi *fNIRS-base*d hanno dimostrato tassi di accuratezza compresi tra il 70% e l'80% nella classificazione delle risposte "sì"/"no", suggerendo la presenza di un'intenzionalità stabile e la possibilità concreta di una comunicazione volontaria[15].

Tali risultati, ottenuti esclusivamente attraverso l'attività cerebrale rilevata da *fNIRS o ERP*, rappresentano un forte indicatore del mantenimento della coscienza e della capacità di giudizio in soggetti completamente paralizzati.

L'implicazione etico-giuridica di questi risultati è significativa: se il pensiero, in assenza di voce o movimento, può essere formalizzato in un linguaggio computazionale e interpretato come volontà consapevole, si pone la questione della validità giuridica di tale manifestazione. È lecito considerare l'*output* di una *BCI* come espressione di consenso informato, di volontà negoziale o di testamento biologico? In che modo si possono soddisfare i requisiti giuridici di volontarietà, comprensione e consapevolezza, in un contesto interamente mediato da dispositivi tecnologici?

Il diritto tradizionale ha storicamente ancorato la validità della manifestazione di volontà a parametri esterni, formali e osservabili, ritenuti essenziali per garantire certezza, tutela della parte debole e prevenzione di abusi. Elementi come la coscienza del soggetto agente, la sua intenzionalità, la comprensione del contenuto e delle conseguenze dell'atto, nonché l'assenza di coercizione o condizionamento sono stati considerati i pilastri della validità negoziale e dell'esercizio di diritti fondamentali. Tali requisiti, che trovano fondamento nei principi generali del diritto civile (si pensi agli artt. 1321, 1322, 1325 e 1427 ss. c.c. in Italia) e nella teoria generale dell'atto giuridico, sono da sempre interpretati alla luce della possibilità di verificare esternamente la volontà del soggetto, attraverso manifestazioni tangibili, scritte, orali o gestuali. Come sottolinea Bianca, il diritto richiede che la volontà si esprima in modo riconoscibile, inequivoco e controllabile, affinché possa essere oggetto di imputazione giuridica e produca effetti normativi certi[16].

Tale impostazione si riflette anche nell'ambito del consenso in ambito

[14] Z.R. Lugo, J. Rodriguez, A. Lechner, *et al.*, "A Vibrotactile P300-Based Brain–Computer Interface for Consciousness Detection and Communication", in *Clinical EEG and Neuroscience*, 45 (2014), n. 1, pp. 14-21.

[15] Si vedano: G. Gallegos-Ayala, A. Furdea, K. Takano, C.A. Ruf, H. Flor, N. Birbaumer, "Brain communication in a completely locked-in patient using functional near-infrared spectroscopy", in *Neurorehabilitation and Neural Repair*, 28 (2014), n. 4, pp. 386–390; S. Rezvani, S. Hosseini-Zahraei, A. Tootchi, *et al*., *op. cit.*

[16] C.M. Bianca, *Diritto civile, vol. III - Il contratto*, ediz. III, Giuffrè, Milano, 2019.

sanitario, dove la legge 219/2017[17] e la giurisprudenza costituzionale[18] ribadiscono la necessità che il consenso sia libero, attuale, consapevole e informato, e che venga espresso in modo non equivoco dal paziente, proprio per garantire il rispetto della sua autodeterminazione e la responsabilità dell'operatore sanitario. La stessa logica si ritrova, più in generale, nella dottrina dell'atto dispositivo: ogni atto giuridico volontario presuppone una soggettività capace di deliberare, decidere e rappresentare la propria intenzione in una forma riconoscibile, anche ai fini della prova e della contestazione[19].

L'introduzione di forme di comunicazione alternative e mediate dalla tecnologia – come avviene nel caso delle *BCI* – solleva una questione centrale per il diritto: è possibile rinunciare alla verificabilità diretta della volontà, affidandosi a una sua traduzione tecnico-algoritmica? E, in caso affermativo, a quali condizioni tale manifestazione può essere considerata giuridicamente equivalente – e non soltanto sostitutiva – rispetto a quella tradizionale? L'assenza di un'espressione corporea autonoma obbliga a riconsiderare il concetto stesso di "manifestazione", non più intesa come mera esteriorizzazione sensibile, ma come riconoscibilità intersoggettiva, anche quando mediata da strumenti neurotecnologici.

In ambito bioetico, questa problematica investe il cuore stesso del concetto di autodeterminazione personale. È moralmente e giuridicamente giustificato escludere un soggetto cosciente dalla possibilità di compiere scelte autonome, solo perché il suo mezzo espressivo non rientra nei canoni canonici della comunicazione? La risposta è negativa: la validità morale dell'autonomia risiede nell'intenzionalità dell'agente, non nel mezzo attraverso cui essa si manifesta[20].

La sfida è dunque duplice: da un lato, occorre riconfigurare il concetto giuridico di manifestazione di volontà, ampliandolo per includere anche modalità espressive non convenzionali; dall'altro, è necessario preservare le garanzie fondamentali – autenticità, consapevolezza e riferibilità personale – che assicurano la validità dell'atto.

Affinché una volontà espressa attraverso segnali neurali possa essere riconosciuta come giuridicamente rilevante, è quindi indispensabile che essa sia raccolta e interpretata all'interno di un quadro procedurale preciso, fondato su protocolli standardizzati, una verifica tecnica affidabile e una supervisione medico-legale indipendente. In questo senso, la sola possibilità tecnica di acquisire un output cerebrale non basta: è necessario che tale espressione sia inserita in un processo che consenta di ricondurla con certezza all'intenzionalità del soggetto.

Proprio per questo, assume un ruolo cruciale la presenza di un meccanismo

[17] Legge 22 dicembre 2017, n. 219, *Norme in materia di consenso informato e di disposizioni anticipate di trattamento*, pubblicata in Gazzetta Ufficiale n. 12 del 16 gennaio 2018.
[18] Si vedano: Corte cost., sentenza 144/2019, in Gazzetta Ufficiale, Serie Generale n. 22 del 26 gennaio 2019; Corte cost., sentenza 242/2019, in Gazzetta Ufficiale, Serie Generale n. 45 del 22 febbraio 2020.
[19] C.M. Bianca, *op. cit*.
[20] O. O'Neill, *Autonomy and trust in bioethics*, Cambridge University Press, Cambridge, 2002.

di supervisione umana che non sia meramente simbolico, ma effettivamente capace di garantire il controllo, la correttezza e la validazione dell'intero processo.

Come si approfondirà nei paragrafi successivi, il Regolamento (UE) 2024/1689 impone, per i sistemi ad alto rischio come le *BCI*, un controllo umano significativo che sia in grado di incidere concretamente sul funzionamento del sistema, intervenendo – ove necessario – per prevenirne gli effetti distorsivi o non intenzionali.

In questo scenario, la supervisione umana non è solo una misura tecnica di accompagnamento, ma diventa lo strumento giuridico e bioetico attraverso cui si può riconoscere come valida una manifestazione di volontà che non si esprime con la voce o la scrittura, ma con pattern neurali codificati. Solo in presenza di un controllo umano strutturato, trasparente e proporzionato ai rischi, si potrà legittimare l'uso delle *BCI* come strumenti attraverso cui esercitare diritti fondamentali, esprimere decisioni rilevanti o compiere atti dispositivi giuridicamente efficaci.

2. Dare voce alla volontà: il riconoscimento giuridico dell'intenzionalità espressa tramite *BCI*

Se l'ordinamento giuridico si fonda, tra i suoi presupposti più elementari, sul principio per cui la volontà deve potersi manifestare in modo riconoscibile e comunicabile, l'esperienza delle persone non comunicanti – in particolare quelle affette da sindromi paralitiche totali (*LIS* e *CLIS*) – obbliga a ripensare criticamente tale assioma. Il paradigma classico della manifestazione della volontà, intesa come esteriorizzazione di un'intenzionalità cosciente attraverso parole, gesti o comportamenti concludenti, entra in crisi di fronte a soggetti che, pur conservando integre le proprie capacità cognitive, sono privi di qualsiasi canale espressivo convenzionale. In questo scenario, il diritto è chiamato a compiere un salto concettuale, ponendosi non più il solo problema della *forma* dell'espressione, bensì quello della *sorgente intenzionale* dell'atto, anche laddove tale volontà si traduca unicamente in un segnale neurale riconosciuto e validato da un dispositivo *BCI*.

Al riguardo, la dottrina ha elaborato una qualificazione concettualmente innovativa: quella degli atti "immoti", per indicare quegli atti giuridicamente rilevanti nei quali la volontà del soggetto non si estrinseca né attraverso la parola né tramite gesti fisicamente percepibili[21]. Tali atti si distinguono dai cosiddetti "atti muti" – che, pur privi di linguaggio verbale, si fondano su un comportamento concludente – in quanto totalmente privi di motilità esterna[22]. Gli atti neuronali,

[21] F.G. Pizzetti, "Libertà di autodeterminazione e protezione del malato nel 'Brain-Computer interfacing': un nuovo ruolo per l'amministratore di sostegno?", in *Rivista Critica del Diritto Privato*, 29 (2011), n. 1, pp. 31-59.
[22] R. Sacco, "Il diritto muto: neuroscienze, conoscenza tacita, valori condivisi", il Mulino, Bologna, 2015.

così, non sono soltanto muti, ma "immoti e non segnici", fondandosi esclusivamente sulla trasduzione di una sequenza di attività cerebrale in un output linguistico tramite interfacce neurali[23].

In questo senso, la nozione di atto giuridico viene svincolata dalla dimensione corporea e ricondotta alla sua essenza più profonda: la volontà intenzionale come fondamento dell'agire giuridico.

Come osservato in dottrina, essi non devono essere intesi come una mera declinazione delle forme note di manifestazione della volontà, ma rappresentano una modalità nuova, tecnicamente mediata, in cui il pensiero – nella sua forma più pura – diventa linguaggio, e il linguaggio diventa atto[24]. Questa trasformazione – da attività sinaptica a produzione semantica riconoscibile – non è solo un passaggio tecnico, ma un nodo strategico per la tenuta dei concetti giuridici di capacità, consenso, rappresentanza e soggettività.

Si tratta, in definitiva, di una "rivoluzione silenziosa" del diritto, che ridefinisce i confini della soggettività giuridica non più in base alla capacità di agire nel mondo fisico, ma alla persistenza di un'intenzionalità comunicabile attraverso mezzi non convenzionali. In questo quadro si inserisce il vivace dibattito sull'attualizzazione del concetto di *comunicazione non verbale*, reso necessario dalle più recenti innovazioni tecnologiche nel campo dell'interazione neurale, le quali offrono oggi a persone affette da SLA – prive di canali espressivi convenzionali ma con capacità cognitive intatte – sistemi alternativi in grado di sfruttare l'unico movimento residuo, spesso quello oculare, per consentire loro l'esercizio diretto dei diritti fondamentali di cui restano titolari.

Ne emerge un'esigenza di giustizia comunicativa: se l'ordinamento pretende di essere inclusivo, deve accettare che anche il silenzio, quando decodificato correttamente, possa diventare parola giuridicamente vincolante.

Invero, già dal 2007, la giurisprudenza di merito italiana ha iniziato, con prudenza ma con crescente attenzione, ad affrontare il tema della rilevanza giuridica della volontà espressa da persone con disabilità fisico-motoria mediante tecnologie assistive, tra cui anche le *BCI* e i sistemi di comunicazione a puntamento oculare combinati con sintetizzatori vocali. In una decisione particolarmente significativa, un giudice tutelare si è pronunciato sul caso di un soggetto affetto da SLA in fase avanzata, pienamente capace di intendere e di volere ma totalmente privo della capacità espressiva tradizionale, che comunicava con il solo movimento oculare attraverso un pannello elettronico e un sintetizzatore vocale. Il tribunale ha respinto la richiesta, avanzata dalla moglie, di nomina quale amministratrice di sostegno, ritenendo che l'individuo fosse in grado di autodeterminarsi e di manifestare

[23] L. Tafaro, "Brain Computer Interface e rilevanza giuridica degli atti 'senza movimento': atti neuronali e *smart working*", in *Il Lavoro nella Giurisprudenza*, (2021), n. 4, pp. 353 ss.
[24] F.G. Pizzetti, *op. cit.*

validamente la propria volontà tramite gli strumenti tecnologici in uso[25].

Il giudice ha fondato il proprio ragionamento sull'assunto per cui il soggetto, pur privo di una capacità materiale di espressione verbale o gestuale, conserva pienamente – ai sensi degli artt. 2, 13 e 32 della Costituzione – il diritto all'autodeterminazione personale. La nomina di un amministratore di sostegno, anche se limitata alla funzione rappresentativa, avrebbe comportato una surrettizia espropriazione della titolarità decisionale, traducendosi in una forma di compressione non proporzionata del diritto alla soggettività giuridica del beneficiario. Si è ritenuto, pertanto, che la manifestazione di volontà così raccolta potesse considerarsi giuridicamente valida, a condizione che fosse attestata da due testimoni maggiorenni.

Tale impostazione evidenzia un'evoluzione importante nella concezione giuridica della capacità, che non è più legata alla *performance* espressiva del soggetto, ma al nucleo intenzionale che la fonda. È la persistenza della coscienza intenzionale, resa tangibile da strumenti tecnologici affidabili, a legittimare il riconoscimento giuridico della volontà.

Un'ulteriore e significativa conferma della possibilità di attribuire rilevanza giuridica alla volontà espressa tramite tecnologie assistive, e in particolare mediante strumenti *BCI*, si rinviene in materia testamentaria. In una pronuncia emblematica, il giudice ha ritenuto giuridicamente valido il testamento di un soggetto affetto da disabilità motoria totale, pienamente lucido e capace di intendere e di volere, che comunicava unicamente tramite puntatore oculare e sintetizzatore vocale. È stato nominato un amministratore di sostegno, non con finalità sostitutive, bensì con il compito di raccogliere, trascrivere e formalizzare la volontà testamentaria espressa dal disponente mediante tali strumenti tecnologici. Il tribunale ha sottolineato che il diritto di disporre per testamento, in quanto espressione della personalità del soggetto e forma estrema di autodeterminazione *post mortem*, non può essere subordinato a requisiti formali rigidi quali la capacità fisica di scrivere, ma deve fondarsi sulla sussistenza della piena capacità naturale, intellettiva e volitiva, coerentemente con gli artt. 2 e 3 Cost[26].

In particolare, è stato riconosciuto – sia in giurisprudenza che in sede notarile – l'esistenza di un vero e proprio diritto alla comunicazione non verbale mediante puntatore oculare, quale strumento idoneo a garantire l'esercizio diretto dei diritti fondamentali, incluso quello di manifestare validamente la propria volontà davanti al notaio, anche in assenza di espressione verbale o gestuale[27]. Tale diritto si configura come una proiezione concreta del principio di uguaglianza sostanziale, trovando un solido fondamento normativo nell'articolo 12 della Convenzione delle

[25] Trib. Sassari, decr., 14 luglio 2007, in *Giurisprudenza di merito*, (2008), n. 5, pp. 1260 ss., e in *Foro Italiano*, I, (2007), p. 3025. Sul tema, v. F. Mazza Galanti, "Il sintetizzatore vocale e la manifestazione di volontà del malato", in *Giurisprudenza di merito*, (2008), pp. 1264 ss.
[26] Trib. Varese, decr., 12 marzo 2012, in *Giurisprudenza di merito*, (2012), pp. 2282 ss.
[27] Trib. Milano, decr., 26 febbraio 2015, n. 11965; Trib. Milano, decr., 26 febbraio 2015, n. 11965.

Nazioni Unite sui diritti delle persone con disabilità (CRPD)[28], al quale riconosce alle persone con disabilità il diritto alla capacità giuridica "su base di uguaglianza con gli altri" e impone agli Stati il dovere positivo di garantire i supporti necessari per l'esercizio effettivo di tale capacità[29].

In questa direzione si colloca l'interpretazione evolutiva offerta dal Consiglio Notarile di Milano nel 2023, secondo cui gli artt. 56 e 57 della l. 16 febbraio 1913, n. 89, che regolano la necessaria presenza dell'interprete nel caso di mutezza o sordomutismo, non possono essere applicati analogicamente a persone che, pur non potendo parlare o gesticolare, sono perfettamente in grado di esprimere la propria volontà attraverso strumenti tecnologici come l'*eye-tracking*[30]. Tali dispositivi, infatti, non richiedono la mediazione linguistica di terzi, ma traducono direttamente l'intenzionalità del soggetto in un linguaggio comprensibile e verificabile, preservando il carattere personale e immediato dell'atto[31].

L'adozione di questa prospettiva ermeneutica, in linea con il principio personalista dell'articolo 2 Cost. e con il dovere di rimozione degli ostacoli previsto dall'articolo 3, comma 2, Cost., rappresenta una svolta culturale e giuridica nel riconoscimento della soggettività giuridica dei soggetti disabili anche nei momenti più delicati e simbolici della loro esistenza giuridica, come la disposizione testamentaria.

Coerentemente con questo approccio, in una pronuncia relativa alla stipula di un contratto di mutuo, il giudice ha riconosciuto piena validità al consenso espresso

[28] Nazioni Unite, "Convenzione sui diritti delle persone con disabilità", adottata il 13 dicembre 2006, firmata dall'Italia il 30 marzo 2007 e ratificata con Legge n. 18 del 3 marzo 2009. Articolo 12 – "Uguale riconoscimento davanti alla legge: Gli Stati Parti riaffermano che le persone con disabilità hanno il diritto di essere riconosciute come persone davanti alla legge e godono della capacità giuridica su base di uguaglianza con gli altri in tutti gli aspetti della vita. Gli Stati devono adottare misure appropriate per fornire accesso al sostegno necessario affinché le persone con disabilità possano esercitare la propria capacità giuridica".

[29] Come chiarito dal Commento Generale n. 1 del Comitato ONU sui diritti delle persone con disabilità, l'esercizio della capacità giuridica non può essere subordinato a criteri cognitivi o comunicativi tradizionali, ma deve essere garantito anche quando si realizza in modalità tecnologicamente assistita. La transizione dal paradigma sostitutivo a quello del *decision-making supportato* impone al diritto di accogliere come valide anche le modalità innovative di manifestazione della volontà, purché affidabili, verificabili e aderenti al principio di autodeterminazione. Comitato ONU sui diritti delle persone con disabilità (2014), *Commento Generale n. 1 – Uguale riconoscimento davanti alla legge (articolo 12 della Convenzione)*, GE.14-03120, adottato l'11 aprile 2014 e pubblicato il 19 maggio 2014.

[30] Gli strumenti di *eye-tracking* (o *eye tracker*) sono dispositivi tecnologici in grado di rilevare e registrare in tempo reale i movimenti oculari di un soggetto, tracciandone la direzione dello sguardo e consentendo l'interazione con interfacce digitali attraverso la sola fissazione visiva. Tali strumenti, impiegati in ambito clinico, ergonomico e comunicativo, possono tradurre la sequenza di movimenti oculari in comandi comprensibili da software di sintesi vocale o scrittura, risultando fondamentali per garantire la comunicazione in soggetti affetti da disabilità motorie gravi.

[31] Consiglio Nazionale del Notariato, *Studio n. 174-2022/P – Sul superamento degli articoli 56 e 57 della legge notarile per i malati di Sindrome Laterale Amiotrofica (SLA)*, a cura di Serena Metallo, 2023.

da una persona con SLA attraverso puntatore oculare e sintetizzatore vocale, sostenendo che l'esercizio personale del diritto alla firma doveva essere garantito mediante firma digitale azionata da *eye-tracker*, espressione di una piena capacità dispositiva, sebbene in forma alternativa[32].

Tuttavia, la giurisprudenza non si è espressa in modo univoco. In un'altra decisione – relativa a un soggetto anch'esso affetto da SLA, ma legalmente e naturalmente capace – il giudice ha invece disposto la nomina di un amministratore di sostegno incaricato di esprimere la volontà di interruzione del trattamento sanitario (nello specifico, la disconnessione dal respiratore artificiale) al posto del beneficiario[33]. Questo scarto ermeneutico mette in luce l'attuale incertezza sistemica nel riconoscimento giuridico della volontà espressa tramite tecnologie assistive, rivelando l'urgenza di un intervento normativo o quantomeno di un consolidamento giurisprudenziale che disciplini, con coerenza e prevedibilità, i presupposti di imputabilità soggettiva della comunicazione neurale.

Alla luce di tali oscillazioni, appare necessario ripensare i criteri tradizionali di riconoscimento della capacità espressiva, alla luce non solo della normativa costituzionale e internazionale in materia di disabilità, ma anche delle più recenti riflessioni bioetiche sul concetto di autodeterminazione tecnologicamente mediata.

In ogni caso, si va delineando un principio di fondo: la disabilità fisica, in sé considerata, non può giustificare alcuna esclusione dalla piena titolarità dei diritti fondamentali, né legittimare surrettizi meccanismi di sostituzione della volontà, ove questa sia esprimibile mediante strumenti tecnologici adeguati. Le interfacce neurali e i dispositivi di comunicazione assistita, ove validati da protocolli affidabili, si configurano sempre più come canali legittimi di esercizio della capacità giuridica, a condizione che siano garantiti i requisiti di autenticità, riferibilità personale, e consapevolezza dell'atto compiuto.

2.1. C*LIS, BCI* e legalità algoritmica: dalla volontà immota alla trasparenza computazionale

La questione si complica in modo decisivo quando si passa dai sistemi BCI supportati da un residuo motorio – come il puntamento oculare – alle interfacce neurali utilizzate da soggetti in condizione di CLIS, totalmente privi di canali espressivi esterni, anche minimi, ma dotati di capacità cognitive intatte. In questo scenario estremo, l'interfaccia neurale non si limita ad amplificare un gesto o a decodificare un segnale visibile: essa si sostituisce integralmente al corpo, generando un codice espressivo interamente computazionale, fondato su *pattern* cerebrali grezzi tradotti in comandi attraverso processi algoritmici ad alta complessità.

[32] Trib. Venezia, Sez. II, decr., 11 aprile 2017, n. 3150.
[33] Trib. Cagliari, decr., 16 luglio 2016, in *Foro it*.

In questo contesto, il pensiero si traduce in atto giuridico senza transitare attraverso alcun veicolo corporeo, con la diretta conseguenza che il fondamento dell'imputabilità soggettiva risulta tecnicamente mediato. La decodifica della volontà avviene infatti per il tramite di un sistema algoritmico che non si limita a registrare un comportamento, ma opera una selezione, un'interpretazione e una traslazione semantica del segnale cerebrale in linguaggio computabile. Ma la rilevanza giuridica attribuibile a questi *output* presuppone che l'intero processo di significazione – dall'attività cerebrale alla formulazione dell'atto – venga sottoposto a un'operazione tecnologica di mediazione profonda, la cui affidabilità è tutt'altro che garantita.

Infatti, come la letteratura neuroscientifica ha ampiamente dimostrato, la decodifica dell'attività neurale è tutt'oggi un'operazione fragile e incerta, condizionata da molteplici variabili fisiologiche e cognitive. La cosiddetta *brain signal variability*, ovvero la variabilità inter- e intra-soggettiva dei segnali cerebrali, rende difficile l'individuazione di *pattern* stabili e univocamente interpretabili[34]. Parametri come età, stress, attenzione o struttura anatomica influenzano significativamente l'affidabilità del segnale. Inoltre, tra il 15% e il 30% degli individui non riescono a generare segnali utilizzabili dalle BCI, una condizione nota come BCI *illiteracy*, le cui cause restano ancora in parte ignote[35].

A queste limitazioni strutturali si aggiungono quelle tecniche, che investono direttamente la possibilità di attribuire effetti giuridici a un pensiero non solo immoto, ma anche semanticamente opaco. Le interfacce neurali fanno uso di algoritmi di *machine learning* altamente sofisticati per riconoscere schemi nei segnali neurali. Tuttavia, la natura dinamica, adattiva e spesso opaca di questi sistemi solleva dubbi radicali: su quali basi è possibile sostenere che l'output algoritmico rifletta autenticamente la volontà dell'individuo?

In assenza di un canale sensoriale esterno di riscontro – visivo, gestuale o linguistico – la validità giuridica dell'atto prodotto si fonda esclusivamente sulla fiducia epistemica nella catena computazionale che ne ha generato il contenuto. Ma tale fiducia è fragile, poiché gli algoritmi stessi sono soggetti a continue riconfigurazioni – dovute all'apprendimento automatico, alla qualità del segnale neurale o alla variabilità neurofisiologica del soggetto – che rendono il processo potenzialmente instabile e non replicabile.

In questo scenario, l'atto "immoto" corre il rischio di divenire un atto "inaccessibile alla verifica": un'espressione giuridicamente rilevante, ma priva di

[34] S. Ryali, Y. Zhang, C. De Los Angeles, K. Supekar, V. Menon, "Deep learning models reveal replicable, generalizable, and behaviorally relevant sex differences in human functional brain organization", in *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 121 (2024), n. 9.

[35] C. Vidaurre, B. Blankertz, "Towards a Cure for BCI Illiteracy", in *Brain Topography*, (2010), pp. 194-198, disponibile su: https://doi.org/10.1007/s10548-009-0121-6. Cfr. M.C. Thompson, "Critiquing the Concept of BCI Illiteracy", in *Science and Engineering Ethics*, 25 (2019), pp. 1217-1233, disponibile su: https://doi.org/10.1007/s11948-018-0061-1.

riscontri intersoggettivi, sottratta alla possibilità di autentica validazione, di contestazione in sede probatoria e di attribuzione responsabile della volontà.

In assenza di un canale comunicativo tangibile, verificabile e condiviso, viene meno la condizione minima di trasparenza necessaria affinché l'intenzione possa essere riconosciuta come fonte legittima di effetti giuridici.

Tale opacità apre una frattura profonda nel rapporto tra volontà e diritto: attribuire rilevanza giuridica a un atto che non transita per alcun mezzo espressivo verificabile significa potenzialmente dissolvere i presupposti stessi della soggettività giuridica e dell'imputabilità. In questo contesto, la volontà non è più manifestata, ma tecnicamente derivata e semanticamente costruita da un sistema algoritmico che opera per approssimazioni probabilistiche, seguendo logiche computazionali che possono incorporare distorsioni, *bias* strutturali o criteri estranei all'identità della persona[36]. Il rischio è che il diritto finisca per riconoscere come autentico ciò che è solo verosimile.

È a partire da questa fragilità epistemica e giuridica che emerge con forza la necessità di ripensare il principio di legalità in chiave algoritmica. La legalità algoritmica[37], infatti, non si limita a verificare la mera conformità normativa dei sistemi, ma si configura come un paradigma regolativo capace di presidiare l'intero processo computazionale attraverso il quale una decisione – o, nel caso delle *BCI*, una manifestazione volontaria – prende forma. Essa si colloca al crocevia tra innovazione tecnologica, diritti fondamentali e legittimazione democratica, e mira a garantire che gli *output* prodotti da sistemi intelligenti siano sempre tracciabili, comprensibili, contestualizzati e sottoposti a un effettivo controllo umano[38].

In questa prospettiva, il principio di legalità algoritmica impone che ogni fase – dalla progettazione degli algoritmi alla raccolta del segnale neurale, dalla sua elaborazione sino alla generazione dell'output giuridicamente rilevante – sia strutturata secondo criteri di trasparenza strutturale, responsabilità tracciabile e auditabilità continua. Solo a tali condizioni è possibile affermare che un atto generato da una *BCI* possa essere considerato, nonostante la sua origine extracorporea e computazionale, conforme ai requisiti di autenticità, imputabilità e riconoscibilità propri di un ordinamento fondato sullo Stato di diritto.

È precisamente in questa direzione che si orienta il Regolamento (UE) 2024/1689 sull'intelligenza artificiale (*AI Act*), il quale riconosce implicitamente le interfacce cervello-computer come sistemi ad alto rischio.

Tale qualificazione non si limita a definire un perimetro di rischio

[36] Si veda: J. Kroll, J. Huey, S. Barocas, *et al.*, "Accountable Algorithms", in *University of Pennsylvania Law Review*, 165 (2017), p. 633 ss., spec. p. 640, nota 14.

[37] Si vedano: A.E. Waldman, "Algorithmic Legitimacy", in W. Barfield (ed.), *The Cambridge Handbook of the Law of Algorithms*, Cambridge University Press, Cambridge, 2020; E. Carloni, "I principi della legalità algoritmica. Le decisioni automatizzate di fronte al giudice amministrativo", in *Rivista di Diritto Amministrativo*, (2020).

[38] Si veda: A. Del Pizzo, "L'informazione neurale accessibile e la crisi dei principi di diritto democratici", in *i-lex*, 17 (2024), n. 2, pp. 11-22.

tecnologico, ma richiama una responsabilità giuridica multilivello, che coinvolge sviluppatori, fornitori, utilizzatori e istituzioni di controllo. In particolare:

- L'articolo 13 impone obblighi di trasparenza e spiegabilità, con l'obiettivo di rendere comprensibili e documentabili i criteri logici e statistici che guidano l'output del sistema, anche in ambiti ad alta complessità computazionale come le BCI;
- L'articolo 14 sancisce il principio della supervisione umana significativa, che diventa decisivo laddove la manifestazione della volontà non è più direttamente accessibile, ma inferita e ricostruita da una macchina. La supervisione non può essere meramente simbolica, ma deve garantire una reale capacità di intervento, correzione e validazione dell'output generato.

Al contempo, il Regolamento richiede valutazioni *ex ante* dei rischi, *audit* periodici e controlli *ex post*, volti a monitorare il comportamento dinamico degli algoritmi, soprattutto in presenza di sistemi che apprendono e si modificano autonomamente nel tempo (*machine learning evolutivo*).

Questo impianto normativo non si esaurisce in un apparato di adempimenti, ma prefigura un modello di legalità sostanziale integrata nei sistemi intelligenti. In tale modello, l'imputazione giuridica di un atto non può prescindere da una ricostruzione verificabile del processo che l'ha generato: è solo attraverso una filiera computazionale *auditabile*, dove ogni passaggio è sottoposto a controllo e verificabilità, che si può ancora parlare, in senso proprio, di volontà giuridicamente rilevante.

3. Il concetto di supervisione umana ai sensi dell'articolo 14

Le riflessioni sviluppate finora hanno messo in luce come l'utilizzo delle BCI in contesti di completa immobilità motoria, specie in condizioni di *Complete Locked-In State* (*CLIS*), sollevi interrogativi inediti sulla riconoscibilità giuridica della volontà individuale, sull'autenticità dell'intenzione e sulla tracciabilità del processo espressivo. L'assenza di una manifestazione corporea accessibile impone al diritto di confrontarsi con forme di comunicazione integralmente algoritmiche, nelle quali la possibilità stessa di attribuire effetti giuridici a un output neurale dipende dalla solidità, dalla trasparenza e dalla verificabilità della catena computazionale che lo ha generato.

In tale scenario, la relazione tra soggetto, volontà e tecnologia diventa talmente stretta da rendere insostenibile qualsiasi approccio meramente formale o tecnocentrico. Occorre piuttosto individuare presìdi normativi e operativi che garantiscano una mediazione umana significativa nel processo di interpretazione e trasduzione dell'intenzione. Il principio di "supervisione umana significativa", codificato all'articolo 14 del Regolamento (UE) 2024/1689, potrebbe rappresentare il punto di convergenza tra esigenze di controllo, istanze democratiche e garanzie di autodeterminazione. In un ecosistema dove l'intenzione è decifrata da algoritmi, la presenza umana deve ritornare a essere garante attiva del processo, non semplice

spettatrice.

È a partire da questa consapevolezza che si apre ora la riflessione sul contenuto e sul significato della supervisione umana nel contesto delle interfacce neurali. Perché se la volontà diventa invisibile, inferita, statistica, allora è solo attraverso un presidio umano capace di leggere, correggere e validare che il diritto potrà ancora riconoscerla come propria.

In un'accezione ampia, il concetto di "supervisione umana" implica il coinvolgimento consapevole e significativo di un essere umano all'interno del processo decisionale algoritmico. Essa può essere realizzata mediante una pluralità di modalità operative, che si differenziano per tempistiche, livelli di intensità e grado di interferenza con il funzionamento autonomo del sistema.

Le diverse configurazioni della supervisione umana vengono distinte e denominate in relazione al momento e alle modalità con cui l'intervento umano si manifesta nel ciclo di vita dell'IA.

Nel contesto della *governance* dei sistemi di intelligenza artificiale, la dottrina e la prassi tecnico-regolatoria distinguono comunemente tra due principali modelli di intervento umano in tempo reale: il paradigma *human-in-the-loop* e quello *human-on-the-loop*[39]. Entrambi i meccanismi presuppongono un'interazione diretta tra l'agente umano e il sistema durante la fase di esecuzione, ma si differenziano in modo significativo quanto al ruolo attribuito all'essere umano. Nel modello *human-in-the-loop*, l'intervento umano costituisce una condizione necessaria affinché l'*output* generato dal sistema produca effetti: ogni decisione automatizzata deve essere previamente validata, approvata o guidata da un operatore umano[40]. Al contrario, nel modello *human-on-the-loop*, l'essere umano assume un ruolo di supervisione passiva e mantiene la facoltà di intervenire nel ciclo decisionale ove necessario, senza che tale intervento sia strutturalmente richiesto.

Oltre a tali configurazioni di supervisione in tempo reale, è possibile individuare ulteriori forme di coinvolgimento umano collocate in fasi differenti rispetto al momento dell'*output*[41]. La prima è rappresentata dalla revisione umana a posteriori, che consente di sottoporre a controllo successivo le decisioni già prodotte dal sistema, con l'eventuale possibilità di annullarle, rettificarle o confermarle *ex post*. La seconda è riconducibile al cosiddetto *human-in-design*, ossia all'inclusione di competenze umane nelle fasi di progettazione, sviluppo, addestramento e validazione del sistema, senza che vi sia un coinvolgimento

[39] R. Koulu, "Human control over automation: EU policy and AI ethics", in *European journal of legal studies*, 12 (2020), n. 1, pp. 9-46.

[40] G. Lazcoz, P. de Hert, "Humans in the GDPR and AIA governance of automated and algorithmic systems. Essential pre-requisites against abdicating responsibilities", in *Computer Law & Security Review*, (2023), pp. 1-31. Sul tema vedi anche: R. Crootof, "A Meaningful Floor for Meaningful Human Control", in *Temple International & Comparative Law Journal*, (2016), pp. 53-54.

[41] T. Enarsson, L. Enqvist, M. Naarttijärvi, "Approaching the human in the loop – legal perspectives on hybrid human/algorithmic decision-making in three contexts", in *Information & Communications Technology Law*, (2022), pp. 123-126.

operativo durante la sua messa in funzione[42].

L'articolo 14 dell'*AI Act* non impone l'adozione di uno specifico modello di supervisione umana, lasciando spazio a un'ampia discrezionalità tecnico-organizzativa nella scelta del meccanismo più adeguato. Ciò che viene espressamente richiesto, tuttavia, è che tale meccanismo sia efficace e commisurato al livello di rischio, al grado di autonomia funzionale del sistema e al contesto concreto in cui esso viene impiegato. Il criterio dell'"efficacia" sembra alludere, nella sostanza, alla necessità che l'intervento umano sia significativo – ovvero dotato di una reale capacità di incidere sull'output del sistema e sul processo decisionale automatizzato. Il riferimento al carattere "significativo" dell'intervento umano si pone in continuità con l'elaborazione dottrinale e interpretativa sviluppatasi in relazione all'articolo 22 del Regolamento generale sulla protezione dei dati (GDPR), e trova altresì eco nelle discussioni relative alla governance dei sistemi d'arma autonomi, dove la *meaningful human control* è divenuta una nozione centrale nel dibattito etico e giuridico[43].

Sebbene l'articolo 14 dell'*AI Act* ponga l'accento sull'efficacia della supervisione umana, esso non definisce in modo tassativo le modalità attraverso cui tale efficacia debba essere raggiunta. Tuttavia, dal tenore stesso della norma possono desumersi alcune esclusioni implicite. In particolare, la disposizione specifica che la supervisione umana debba essere esercitata "durante il periodo in cui [i sistemi] sono in uso", escludendo così la sufficienza di un mero coinvolgimento umano limitato alle fasi di progettazione, sviluppo o addestramento. Ne deriva che meccanismi riconducibili esclusivamente alla logica dello *human-in-design* non possono considerarsi idonei, di per sé, a soddisfare l'obbligo di supervisione delineato dall'articolo 14.

In termini strettamente giuridici, è opportuno distinguere tra l'obbligo di predisporre un meccanismo di supervisione, che incombe sul fornitore del sistema e trova fondamento nell'articolo 14, e l'obbligo di esercitare concretamente tale supervisione, che grava sul soggetto *deployer* e trova il suo fondamento normativo nell'articolo 26, paragrafo 2 *AI Act*. Quest'ultimo impone infatti ai *deployers* l'onere di assicurare che la supervisione umana sia effettivamente assegnata, attribuendo rilievo operativo e organizzativo alla funzione umana nel ciclo di vita del sistema ad alto rischio[44].

[42] R. Crootof, M.E. Kaminski, W. Nicholson, "Humans in the Loop", in *Vanderbilt Law Review*, (2023), pp. 429-510; Sul tema si veda anche M.L. Jones, "The right to a human in the loop: Political constructions of computer automation and personhood", in *Social studies of science*, (2017), pp. 47-78; K. Brennan-Marquez, K. Levy, D. Susser, "Strange Loops: Apparent versus Actual Human Involvement in Automated Decision Making", in *Berkeley Technology Law Journal*, (2019) pp. 745-749.

[43] G. Mecacci, *Research Handbook on Meaningful Human Control of Artificial Intelligence Systems*, Edward Elgar Publishing, Cheltenham, 2024, p. 372.

[44] N.A. Smuha, "The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence", in C*omputer Law Review International*, (2019), pp. 20-97; N.A. Smuha, K. Yeung, "The European Union's AI Act: beyond motherhood and apple pie?", in N.A Smuha (ed.), *The Cambridge*

Rimane oggetto di discussione se l'articolo 14 dell'*AI Act* esiga, nelle tecnologie di *Brain Computer Interfaces*, la predisposizione di una possibilità di intervento umano in tempo reale[45]. Un'indicazione in tal senso può essere tratta dall'articolo 14, paragrafo 4, lettera e), il quale prescrive che la persona incaricata della supervisione umana debba essere in grado di "intervenire nell'operazione […] o interrompere il sistema attraverso un pulsante di arresto o una procedura analoga". Tuttavia, il tenore letterale della disposizione non chiarisce se i requisiti elencati al paragrafo 4, debbano intendersi in senso cumulativo – imponendo cioè il rispetto di tutte le condizioni ivi indicate – oppure in senso alternativo, consentendo l'adozione selettiva delle misure in base al contesto d'uso.

Nel primo scenario interpretativo, la norma configurerebbe un obbligo generalizzato di garantire un meccanismo di intervento umano in tempo reale, riconducibile ai modelli *human-in-the-loop* o *human-on-the-loop*, a prescindere dalla criticità del contesto applicativo. In tal caso, la supervisione tempestiva rappresenterebbe un elemento strutturale ed imprescindibile per ogni sistema di IA basato su tecnologie di BCI. In alternativa, qualora si optasse per una lettura disgiuntiva delle lettere da a) ad e) del paragrafo 4, la necessità di un intervento in tempo reale dovrebbe essere valutata caso per caso, sulla base dei criteri di adeguatezza e proporzionalità rispetto ai rischi concreti, al grado di autonomia tecnica del sistema e all'ambiente operativo. Tale interpretazione sembra più coerente con l'approccio basato sul rischio che informa l'intero impianto regolatorio dell'*AI Act*.

3.1. La supervisione umana nella *governance* dei sistemi di IA basati su tecnologie *BCI*

La supervisione umana non costituisce un fine in sé, ma rappresenta uno strumento funzionale al perseguimento di una pluralità di finalità, la cui rilevanza e configurazione variano sensibilmente in base ai contesti applicativi e agli approcci disciplinari. Pur senza pretese di esaustività, gli obiettivi sottesi alle politiche di supervisione umana possono essere ricondotti, in via sistematica, a due categorie concettuali – parzialmente sovrapposte ma analiticamente distinguibili –: (1) finalità orientate al risultato, che mirano a garantire l'accuratezza, la sicurezza e l'affidabilità delle decisioni generate dai sistemi di IA; e (2) finalità orientate alla responsabilità, che mirano ad assicurare la tracciabilità delle decisioni e l'attribuzione chiara delle responsabilità giuridiche in caso di esiti dannosi o controversi.

Handbook on the Law, Ethics and Policy of Artificial Intelligence, Cambridge University Press, Cambridge, (2025), pp. 132-145.

[45] Cfr. R. Crootof, M.E. Kaminski, W. Nicholson, *op. cit.*, pp. 440-442. Sul tema si veda anche B. Green, "The flaws of policies requiring human oversight of government algorithms", in *Computer Law & Security Review*, (2022), pp. 21-45.

Gli obiettivi orientati all'*output* affidano alla supervisione umana il compito di identificare e correggere risultati imprecisi, erronei o indesiderati generati dai sistemi di intelligenza artificiale. Tali errori possono derivare da malfunzionamenti tecnici, da lacune nei dati di addestramento o, più semplicemente, dalla fallibilità intrinseca dei modelli algoritmici. La logica sottesa a questo approccio risiede nella convinzione che l'intervento umano, fondato su conoscenze pregresse, competenze esperienziali o persino intuizioni non formalizzabili, possa rappresentare un efficace correttivo rispetto ai limiti delle decisioni automatizzate.

Un caso emblematico che viene spesso richiamato per illustrare il valore della supervisione umana nei confronti dell'*output* automatizzato è il cosiddetto "incidente Petrov". Il 26 settembre 1983, Stanislav Petrov, all'epoca tenente colonnello delle forze di difesa aerea sovietiche, decise di non dare seguito all'allarme lanciato da un sistema automatizzato di rilevamento, che segnalava – erroneamente – l'avvio di un attacco nucleare da parte degli Stati Uniti[46]. Basandosi su una valutazione intuitiva e nonostante l'apparente urgenza dell'allerta, Petrov scelse di non attivare la procedura di risposta, evitando così, secondo molte ricostruzioni, un conflitto nucleare su scala globale. Questo episodio, pur collocandosi in un contesto storico e tecnologico ben diverso da quello odierno, continua a essere citato come paradigma del valore insostituibile del giudizio umano in situazioni ad elevatissimo rischio.

La funzione correttiva attribuita all'intervento umano risulta particolarmente rilevante nei casi in cui l'essere umano sia in grado di superare i limiti cognitivi e conoscitivi dell'algoritmo[47]. Le carenze strutturali degli attuali sistemi di intelligenza artificiale – in particolare, la loro incapacità di cogliere pienamente le sfumature del contesto sociale, culturale o morale – possono condurre a risultati distorti, discriminatori o, più semplicemente, privi di significato nel contesto reale in cui vengono applicati. In tali circostanze, la supervisione umana può intervenire in modo decisivo, introducendo elementi di giudizio valoriale, sensibilità etica o conoscenze contestuali che sfuggono al calcolo computazionale[48].

Un esempio emblematico di questa dinamica è descritto da Margot Kaminski: in un caso documentato, un sistema algoritmico impiegato per il rilevamento di frodi nei programmi di assistenza alimentare ha segnalato come sospetti diversi negozi di alimentari di proprietà somala a Seattle, sulla base di "transazioni atipiche". Tali anomalie, tuttavia, trovavano spiegazione plausibile nei modelli di consumo specifici della comunità di immigrati dell'Africa orientale insediata in

[46] Questo episodio è descritto da Kiel Brennan-Marquez e Stephen Henderson nell'articolo "Artificial Intelligence and Role-Reversible Judgment", in *Journal of Criminal Law and Criminology*, (2019), pp. 137-146.

[47] D. Solove, H. Matsumi, "AI, Algorithms, and Awful Humans", in *Fordham Law Review*, (2024), pp. 19-23.

[48] M. Hildebrandt, "Privacy As Protection of the Incomputable Self: Agonistic Machine Learning", in *Theoretical Inquiries in Law*, (2019), pp. 20-83; R. Binns, "Human Judgment in algorithmic loops: individual justice and automated decision-making", in *Regulation & Governance*, (2022), pp. 185-197.

quell'area, legati a pratiche collettive di acquisto e redistribuzione. In questo contesto, un operatore umano dotato di consapevolezza sociale e culturale avrebbe potuto rilevare la non corrispondenza tra il sospetto generato dal sistema e la realtà fattuale, correggendo così un'ingiustizia latente e potenzialmente dannosa.

Gli obiettivi orientati alla responsabilità, invece, si fondano sull'assunto che le macchine non possano, in senso proprio, essere titolari di responsabilità giuridica o morale. La presenza di un essere umano nel ciclo decisionale o operativo dei sistemi di intelligenza artificiale *BCI* risponde, in questo quadro, all'esigenza di individuare un soggetto imputabile – e quindi giuridicamente responsabile – per le conseguenze delle azioni automatizzate[49].

Come ha evidenziato Madeleine Elish, tuttavia, la supervisione umana finisce spesso per svolgere una funzione di "assorbimento" della responsabilità (*moral crumple zone*), che va ben oltre il controllo effettivamente esercitabile dall'operatore sul comportamento del sistema. In tali casi, il sorvegliante umano viene configurato come terminale della responsabilità, anche qualora l'esito dannoso sia dipeso da un malfunzionamento tecnico, da scelte progettuali opache o da una distribuzione ambigua delle competenze tra sviluppatori, fornitori e utilizzatori del sistema di IA.

La funzione a discarico della supervisione umana, che tende a deresponsabilizzare l'infrastruttura tecnico-organizzativa a vantaggio dell'attribuzione della colpa a un singolo individuo, può essere oggetto di un uso strategico e opaco da parte dei fornitori di sistemi complessi. Come evidenziato nel caso ampiamente discusso in dottrina dei veicoli a guida autonoma di Tesla, il trasferimento del controllo al conducente umano pochi istanti prima dell'impatto servirebbe a costruire, *ex post*, una narrazione di interruzione dell'automazione, funzionale a ricondurre la responsabilità all'utente[50].

Tale dinamica, se applicata a tecnologie *BCI* impiegate da soggetti affetti da patologie altamente invalidanti – come nel caso di pazienti affetti da SLA – porrebbe numerosi e delicati interrogativi. In questi scenari, il sistema *BCI* non rappresenta un mero supporto, ma costituisce la sola interfaccia attraverso cui l'individuo può esprimere volontà e agire nel mondo esterno. L'idea stessa di "trasferimento del controllo" all'utente umano, in assenza della tecnologia, risulterebbe priva di significato operativo: il soggetto, di fatto, non sarebbe in condizione di esercitare alcuna azione autonoma, né tanto meno di assumere il controllo diretto del sistema.

In un simile contesto, un'eventuale strategia volta a scaricare sul paziente la responsabilità per un malfunzionamento del sistema *BCI* sarebbe giuridicamente

[49] Cfr. R. Crootof, M. E. Kaminski, W. Nicholson, *op. cit.*, pp. 482-483; J. Beck, T. Burri, "From human control in international law to human oversight in the new EU act on artificial intelligence", in G. Mecacci (ed.), *Research Handbook on Meaningful Human Control of Artificial Intelligence Systems*, Edward Elgar Publishing, Cheltenham, 2024, pp.107-109.
[50] A. Bibals, "Legal requirements on explainability in machine learning", in *Artificial Intelligence and Law*, (2021), pp. 136-149.

inaccettabile e moralmente insostenibile. Essa determinerebbe una distorsione grave del principio di imputazione soggettiva, violando i presupposti minimi di capacità, volontarietà e consapevolezza richiesti per attribuire responsabilità – specie se penale o risarcitoria – a un individuo.

Inoltre, configurerebbe una violazione manifesta dei diritti fondamentali della persona disabile, in particolare del principio di uguaglianza sostanziale, della dignità umana e del diritto a misure assistive adeguate.

3.2. Il principio della supervisione umana nell'*AI Act*

Il ruolo attribuito al sorvegliante umano nel contesto del Regolamento (UE) 2024/1689 emerge con particolare chiarezza dall'articolo 14, paragrafo 2, il quale stabilisce che il controllo umano deve "mirare a prevenire o ridurre al minimo i rischi per la salute, la sicurezza o i diritti fondamentali". Tale formulazione evidenzia una chiara finalizzazione *outcome-oriented* della supervisione: il sorvegliante umano non è chiamato a garantire la mera presenza simbolica o formale di un controllo antropico, bensì a svolgere una funzione effettivamente protettiva rispetto ai possibili impatti negativi del sistema di intelligenza artificiale.

Questa impostazione trova conferma nella relazione illustrativa che accompagna la proposta della Commissione europea, dove si sottolinea che "gli obblighi in materia di *test ex ante*, gestione del rischio e sorveglianza umana faciliteranno anche il rispetto di altri diritti fondamentali, riducendo al minimo il rischio di decisioni errate o distorte assistite dall'IA". Ne deriva che la supervisione umana, così come delineata dal legislatore europeo, assume il profilo di un presidio funzionale di protezione sostanziale, volto ad assicurare che i sistemi ad alto rischio siano concepiti, implementati e utilizzati in modo conforme ai principi di sicurezza, affidabilità e tutela dei diritti.

Tale prospettiva impone, di conseguenza, una ridefinizione qualitativa del ruolo umano: non si tratta di una funzione meramente ancillare o reattiva rispetto alla tecnologia, ma di una presenza dotata di competenze, autorità e strumenti concreti per incidere sul processo decisionale automatizzato, intervenendo in modo tempestivo, informato e responsabile. In questo senso, il sorvegliante umano diventa un elemento strutturale del sistema di governance dell'intelligenza artificiale, in un'ottica chiaramente orientata alla prevenzione del danno e alla promozione dei valori fondanti dell'ordinamento europeo[51].

Tuttavia, il riferimento espresso ai diritti fondamentali contenuto nell'articolo 14, paragrafo 2, del Regolamento (UE) 2024/1689 suggerisce che la funzione della supervisione umana non si esaurisca in una logica meramente orientata ai risultati, ma includa anche obiettivi propriamente *process-oriented*. In particolare, la

[51] J. Laux, "Institutionalised distrust and human oversight of artificial intelligence: towards a democratic design of AI governance under the European Union AI Act", in *AI & Society*, (2023), n. 1, pp. 69-73.

disposizione richiama un insieme di valori e garanzie che presuppongono non soltanto l'efficacia strumentale della supervisione nel prevenire danni, ma anche la qualità e la legittimità del processo decisionale stesso, in relazione al rispetto dell'autonomia individuale, della trasparenza e della partecipazione.

Tale interpretazione trova fondamento nei documenti strategici dell'Unione Europea che hanno preceduto e accompagnato l'adozione dell'*AI Act*. Tra questi, un rilievo centrale è assunto dagli orientamenti etici elaborati dall'*High-Level Expert Group on Artificial Intelligence* (*HLEG*), nei quali si afferma chiaramente che "gli esseri umani che interagiscono con i sistemi di IA devono essere in grado di mantenere un'autodeterminazione piena ed efficace su sé stessi". Per conseguire tale obiettivo, si raccomanda che l'allocazione delle funzioni tra esseri umani e sistemi automatizzati garantisca "una significativa opportunità di scelta umana", assicurando al tempo stesso una supervisione sostanziale e continua dei processi automatizzati[52].

Analoghe affermazioni si rinvengono nelle comunicazioni ufficiali della Commissione europea, che ha sottolineato come il controllo umano rappresenti un elemento chiave per "garantire che un sistema di IA non comprometta l'autonomia umana". In questa prospettiva, la supervisione umana non svolge soltanto una funzione protettiva *ex post*, ma contribuisce alla costruzione di un rapporto fiduciario tra l'individuo e il sistema tecnologico, promuovendo un contesto di legittimazione democratica, responsabilità istituzionale e rispetto della persona.

Il requisito della supervisione umana, pertanto, si configura come uno strumento polifunzionale, capace di soddisfare contemporaneamente esigenze di sicurezza, salvaguardia dei diritti e promozione di finalità sistemiche più ampie, quali l'inclusività, la trasparenza procedurale e la coesione sociale. In definitiva, l'umanizzazione del processo decisionale algoritmico – anche nei suoi aspetti formali – si pone come condizione necessaria per la piena accettabilità e sostenibilità democratica dell'intelligenza artificiale in ambito europeo[53].

In definitiva, l'articolo 14 del Regolamento (UE) 2024/1689 manifesta una preminente vocazione *outcome-oriented*, ponendo l'accento sulla prevenzione e sulla mitigazione dei rischi per la salute, la sicurezza e i diritti fondamentali. Tuttavia, se interpretato alla luce del più ampio contesto normativo e valoriale che ha condotto all'adozione dell'*AI Act*, tale articolo ingloba anche finalità *process-oriented*, in particolare quelle relative alla salvaguardia dell'autonomia umana, alla promozione dell'azione consapevole e all'incremento della fiducia nei sistemi di intelligenza artificiale[54]. In tal senso, la supervisione umana assume una funzione non meramente tecnica, ma anche relazionale, simbolica e istituzionale.

Ciò che risulta invece significativamente assente tra gli obiettivi espressi

[52] AI HLEG, *Ethics Guidelines for Trustworthy AI*, 8 April 2019.
[53] F. Romeo, "Cooperation, Law and Artificial Intelligence Technologies", in F. Santoianni, G. Gianni, A. Ciasullo (eds.), *Mind, Body and Digital Brains*, Springer, Cham, 2024, pp. 253-263.
[54] F. Romeo, "Normatività: uno studio sulla struttura, funzione e quantificazione", in *L'Ircorcervo*, 22 (2022), n. 2, pp. 59-76.

dell'articolo 14 è una finalizzazione esplicita alla garanzia della responsabilità.

La norma non attribuisce alla supervisione umana una funzione diretta di *accountability*, né stabilisce criteri per l'attribuzione della responsabilità in caso di malfunzionamenti o danni causati da sistemi di IA. Ciononostante, tale assenza non impedisce che il sorvegliante umano finisca, nei fatti, per diventare il terminale della responsabilità, soprattutto in contesti in cui le strutture giuridiche di imputazione restano fortemente antropocentriche.

In altri termini, la responsabilità del sorvegliante umano può emergere non come finalità deliberata dell'architettura regolatoria, ma come effetto collaterale – e potenzialmente distorsivo – dei requisiti di sorveglianza previsti dall'articolo 14. In mancanza di una chiara allocazione normativa delle responsabilità lungo l'intera filiera del sistema di IA (dallo sviluppatore al fornitore, fino all'utilizzatore), l'operatore umano può trovarsi a fungere da *crumple zone* giuridica, assorbendo le conseguenze di decisioni erronee o disfunzionali su cui, in molti casi, non ha esercitato un controllo effettivo o informato[55]. Questo scenario solleva interrogativi rilevanti in ordine alla legittimità dell'imputazione soggettiva della responsabilità, nonché alla necessità di un ripensamento strutturale dei criteri di responsabilità nell'ecosistema dell'intelligenza artificiale.

4. I limiti della supervisione umana e le prospettive regolatorie

Si può affermare, senza esitazione, che il principio della supervisione umana costituisce un elemento fondativo nella disciplina dei sistemi di intelligenza artificiale ad alto rischio, come chiaramente attestato dall'articolo 14 del Regolamento (UE) 2024/1689. Tale previsione normativa è il risultato di una solida elaborazione politico-istituzionale e gode, sotto il profilo regolatorio, di un'elevata affidabilità sistemica[56]. Tuttavia, l'esperienza empirica solleva importanti interrogativi circa l'effettività del coinvolgimento umano nel garantire i risultati attesi.

Diversi studi hanno infatti dimostrato che i sorveglianti umani, pur formalmente presenti nel ciclo operativo dei sistemi di IA, non sempre riescono a identificare suggerimenti errati, distorti o indesiderati prodotti dall'algoritmo[57]. In

[55] F. Corona "L'evoluzione del processo decisionale tra giudizio automatico e sostituzione giudiziaria", in *L'Ircocervo*, 23 (2024), n. 1, pp. 176-193. Sul tema vedi anche F. Corona, M. Ferrarese, "The future without a Judge", in *Rivista Italiana di informatica e diritto*, 25 (2025), n. 1, pp. 310-359.

[56] Si veda sul tema, Aa.Vv., "Intelligenza Artificiale e diritto", a cura di E. Gabrielli, U. Ruffolo, in *Giurisprudenza italiana*, (2019), pp. 1656 ss.; A.C. Amato Mangiameli, "Algoritmi e big data. Dalla carta sulla robotica", in *Rivista di filosofia del diritto*, (2019), n. 1, pp. 117-120; F. Romeo, *Il diritto artificiale*, Giappichelli, Torino, 2002, p. 176; F. Romeo, "Esplorazioni nel diritto artificiale", in *I-Lex*, (2004), n. 1, pp. 37-38.

[57] M. Tegmark, *Vita 3.0. Essere umani nell'era dell'intelligenza artificiale*, Raffaello Cortina Editore, Milano, 2018, p. 143. Sul tema dell'applicazione di algoritmi giuridici, si veda anche: F.

alcuni casi l'intervento umano ha condotto alla correzione di risultati che erano, in realtà, pienamente corretti, con effetti peggiorativi sulla qualità complessiva del processo decisionale. In tali situazioni, il ruolo del sorvegliante umano non soltanto non migliora la performance del sistema, ma finisce per comprometterne l'affidabilità e la coerenza.

Le ragioni alla base di questo fenomeno sono molteplici e includono, da un lato, *bias* cognitivi e psicologici che influenzano il comportamento umano in presenza di sistemi automatizzati – come l'eccessiva deferenza verso l'output della macchina (*automation bias*) o, al contrario, un'irragionevole sfiducia (*algorithm aversion*) – e, dall'altro lato, condizioni operative che limitano le capacità del sorvegliante, quali il sovraccarico informativo, la pressione temporale, o la mancanza di trasparenza del sistema stesso[58].

Occorre riconoscere, pertanto, che non tutte le criticità possono essere affrontate mediante l'introduzione di ulteriori obblighi normativi o formali.

Alcune limitazioni derivano direttamente dai confini strutturali della cognizione umana e non possono essere superate esclusivamente attraverso interventi regolatori. È dunque essenziale che tali vulnerabilità vengano prese in considerazione sin dalla fase di progettazione dei meccanismi di supervisione, orientando la configurazione dei sistemi verso soluzioni che non si limitino a prevedere l'intervento umano, ma lo rendano effettivamente sostenibile, informato e significativo.

Non rientra negli obiettivi di questo contributo affrontare in modo sistematico le proposte di riforma delle politiche di supervisione umana né l'esame delle possibili salvaguardie alternative. Tuttavia, alcune considerazioni di carattere generale possono essere tratte a fini interpretativi e applicativi, in particolare con riferimento all'articolo 14 del Regolamento (UE) 2024/1689.

Anzitutto, occorre riconoscere che la supervisione umana non costituisce una soluzione universale ai rischi derivanti dall'impiego di sistemi di intelligenza artificiale. L'affidamento eccessivo su tale strumento come salvaguardia privilegiata può risultare controproducente, soprattutto laddove serva – esplicitamente o implicitamente – a giustificare un allentamento degli *standard* previsti per altre misure di protezione. Questo rischio diventa particolarmente rilevante se si considera la tendenza, talvolta riscontrabile nella prassi regolatoria o nella progettazione tecnica, a sovrastimare la capacità dell'intervento umano di

Romeo, "Algoritmi giuridici di *machine learning* e controllo del giudicato interculturale", in *Rivista di filosofia del diritto*, (2023), n. 1, pp. 163-165; F. Romeo, "Algoritmi di giustizia ed equità nel diritto", in *I-Lex*, 14, (2021), n. 1, pp. 11-12.

[58] D. Kahneman, A. Tversky, "Judgment under Uncertainty: Heuristics and Biases", in *Science,* 185 (1974), n. 4157, pp. 1124-1131. In questo articolo gli autori hanno introdotto il concetto di euristiche e pregiudizi nel processo decisionale. Sul tema vedi anche A. Chiappetta, "Navigating the AI Frontier: European Parliamentary Insights on Bias and Regulation, Preceding the AI Act", in *Internet Policy Review*, (2023) pp. 35-54.

neutralizzare, in modo isolato, i rischi sistemici e strutturali dell'IA[59].

In tale prospettiva, è fondamentale sottolineare che, nell'ambito del sistema integrato di garanzie previsto per i sistemi di IA ad alto rischio, la previsione di un meccanismo di supervisione umana non dovrebbe mai essere utilizzata come argomento per derogare ad altri requisiti di sicurezza, trasparenza o robustezza tecnica. Ciò è particolarmente rilevante in relazione all'articolo 6, paragrafo 3, lettera c), che, pur in un contesto limitato, consente una mitigazione di altre garanzie qualora sia presente un "adeguato controllo umano". Tale disposizione richiede un'interpretazione restrittiva e un'applicazione prudente, affinché non si trasformi in una clausola di esenzione sistematica o in un varco normativo suscettibile di indebolire il livello complessivo di tutela.

La supervisione umana deve essere intesa come uno degli strumenti – e non l'unico – di un sistema multilivello di protezione, la cui efficacia dipende dal corretto bilanciamento e dall'integrazione coerente tra le diverse salvaguardie previste dal Regolamento. Solo un approccio olistico e non gerarchico ai meccanismi di tutela potrà garantire un'effettiva protezione dei diritti fondamentali e una governance responsabile dell'intelligenza artificiale.

Infine, la relazione tra supervisione umana e responsabilità giuridica si configura come intrinsecamente complessa e ambivalente. I soggetti umani incaricati della sorveglianza rischiano, soprattutto nei sistemi di BCI, di essere investiti di un ruolo puramente simbolico o scarsamente operativo, al punto da trasformarsi in mere "spugne di responsabilità" (*responsibility sponges*), come efficacemente descritto in letteratura. In tali situazioni, i sorveglianti finiscono per essere ritenuti responsabili degli esiti negativi del funzionamento del sistema, pur non disponendo di un margine di azione o di un controllo effettivo che giustifichi, sul piano giuridico o etico, l'attribuzione di tale responsabilità.

Questo scollamento tra *agency* e imputabilità non solo compromette il principio fondamentale di responsabilità personale, ma rischia anche di produrre effetti distorsivi nelle dinamiche organizzative e decisionali.

Una simile dinamica è stata osservata da Ben Green, il quale ha rilevato come, nella prassi amministrativa, l'attuazione delle politiche di supervisione umana finisca per scaricare la responsabilità sui *street-level bureaucrats*, ovvero sugli operatori di front-line che interagiscono direttamente con gli utenti, anche quando le scelte rilevanti sono state prese altrove – ad esempio nella progettazione o nella configurazione algoritmica del sistema[60].

Tuttavia, non si può ignorare il fatto che l'attribuzione di responsabilità a soggetti umani rappresenta una componente essenziale dello Stato di diritto, nonché un presupposto necessario per garantire la giustizia individuale. In assenza di

[59] A. Simoncini, "L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà", in *BioLaw Journal. Rivista di BioDiritto*, (2019), n. 1, pp. 76-82; L. Solum, "Artificially Intelligent Law", in *BioLaw Journal. Rivista di BioDiritto*, (2019), n. 1, pp. 156-169.

[60] B. Green, "The flaws of policies requiring human oversight of government algorithms", in *Computer Law & Security Review*, (2022), pp. 22-45.

riferimenti umani identificabili, i meccanismi di tutela e ricorso si svuoterebbero di significato, e verrebbe meno la possibilità stessa di esercitare diritti nei confronti di decisioni automatizzate.

L'articolo 14 dell'*AI Act* non affronta in modo esplicito la questione, limitandosi a definire i requisiti funzionali della supervisione umana, senza interrogarsi sulla compatibilità di tali requisiti con un'imputazione equa e proporzionata della responsabilità. Tuttavia, è auspicabile che questa problematica venga adeguatamente considerata sia nella progettazione dei meccanismi di supervisione, sia nella definizione dei quadri normativi di responsabilità. Appare necessario superare il modello monodirezionale di attribuzione della responsabilità centrato esclusivamente sul supervisore umano nel punto di utilizzo, a favore di una distribuzione più articolata lungo l'intero ciclo di vita del sistema di IA – dalla progettazione e sviluppo, alla validazione, fino alla fase di *deployment*. Solo attraverso un modello multilivello e sistemico di *accountability* sarà possibile garantire una responsabilità effettiva, proporzionata e conforme ai principi fondamentali dell'ordinamento giuridico[61].

5. Riflessioni conclusive

L'indagine sviluppata ha messo in luce come l'impiego delle *BCI*, specie nei casi di *LIS* e *CLIS*, interroghi in profondità le categorie giuridiche classiche, costringendo il diritto a confrontarsi con forme di espressione della volontà radicalmente nuove. Non si tratta soltanto di comprendere se questi soggetti possano validamente compiere atti giuridici – il che rimanderebbe alla nozione tecnica di capacità di agire – ma di affrontare una questione più radicale: la piena effettività della capacità giuridica.

In senso teorico, ogni persona conserva la capacità giuridica per il solo fatto di esistere come soggetto di diritto. Tuttavia, quando le condizioni comunicative impediscono l'espressione della volontà secondo codici tradizionali, questa capacità rischia di diventare astratta, priva di effettività e non riconosciuta nella prassi. Il diritto, storicamente ancorato alla corporeità dell'espressione – la voce, la scrittura, il gesto – fatica a riconoscere forme di volontà "immota", non segnica, come quelle espresse tramite *pattern* neurali. Ciò comporta un paradosso giuridico: soggetti cognitivamente presenti e giuridicamente capaci, ma di fatto esclusi dalla possibilità di esercitare i propri diritti fondamentali.

È in questo spazio che le *BCI* si inseriscono non come semplici ausili tecnici, ma come dispositivi di riattivazione giuridica, in grado di restituire ai soggetti

[61] F. Corona, "IA e Giustizia", in M. Iaselli (a cura di), *AI ACT: Principi, regole ed applicazioni pratiche del Reg. UE 1689/2024*, Maggioli Editore, Santarcangelo di Romagna, (2024) pp. 155-187; F. Corona, "IA e regolamentazione: una comparazione tra approcci normativi per un equilibrio sostenibile", in U. Comite, A. Kostyuk (a cura di), *Sostenibilità e intelligenza artificiale: resilienza o panacea?*, Wolters Kluwer, Milano, 2025, pp. 101-130.

silenziosi un canale espressivo alternativo. Tuttavia, il ricorso all'intelligenza artificiale e alla computazione algoritmica nella decodifica dell'attività neurale non è neutro né privo di rischi. La variabilità dei segnali cerebrali, l'opacità dei modelli di *machine learning* e l'assenza di canali di verifica intersoggettiva sollevano dubbi significativi sulla autenticità, riferibilità e imputabilità degli atti generati.

In questo contesto, il principio di supervisione umana significativa, previsto dall'articolo 14 dell'*AI Act*, assume una valenza decisiva. Non si tratta di una mera misura tecnica o di un elemento accessorio della *compliance* regolatoria: la supervisione umana è, qui, il presidio giuridico attraverso cui il diritto può continuare a riconoscere soggetti dove altrimenti vedrebbe solo sistemi e segnali. Affinché la volontà computazionalmente espressa possa essere giuridicamente rilevante, è necessario che vi sia un controllo umano effettivo, capace di garantire che ciò che viene codificato come "volontà" corrisponda a una vera intenzionalità consapevole. La supervisione, dunque, è strumento di protezione della persona, ma anche di legittimazione del diritto stesso, che ritrova nella mediazione umana una forma di garanzia della propria razionalità e giustizia.

Questo porta a una riflessione finale: non basta proclamare l'universalità dei diritti fondamentali, se non si è disposti a riconoscerli anche là dove si manifestano in forme nuove, fragili, decentrate. La capacità giuridica non può essere vincolata alla forma, ma va intesa come diritto a essere riconosciuti nella propria soggettività, anche oltre i limiti della corporeità attiva. Il diritto, se vuole restare fedele al suo nucleo personalista e inclusivo, deve sapersi aprire a forme di comunicazione neuroassistita, purché supportate da adeguate garanzie procedurali, standard condivisi e responsabilità distribuite.

In definitiva, solo una *governance* multilivello – capace di integrare profili tecnici, normativi ed etici – potrà assicurare che la soggettività giuridica non venga negata dal silenzio, ma riconosciuta anche quando si esprime attraverso nuove interfacce tra mente, macchina e diritto. È questo il compito che ci attende: ripensare la giustizia anche per chi non può più parlare.