OLTRE IL *RISK-BASED APPROACH*, VERSO UN *"COSTITUZIONALISMO ALGORITMICO SOSTANZIALE"* NELL'ERA DELL'INTELLIGENZA ARTIFICIALE [*]

Beyond the Risk-Based approach, towards a 'substantial algorithmic constitutionalism' in the era of artificial intelligence.

Attilio Luigi Maria Toscano [**]

Abstract (It): il saggio analizza criticamente il quadro normativo europeo sull'Intelligenza Artificiale, con particolare riferimento al Regolamento (UE) 2024/1689 (*AI Act*) e al relativo disegno di legge di adeguamento italiano. Si argomenta che l'architettura giuridica, fondata su una definizione puramente funzionale di IA e su un approccio basato sul rischio (*risk-based approach*), sia strutturalmente inadeguata a garantire una tutela effettiva dei diritti fondamentali. Tale modello, ignorando la questione cruciale dell'opacità algoritmica (*black box*), "*de-costituzionalizza*" i diritti, trasformandoli da principi inviolabili a variabili negoziabili in un calcolo di convenienza, e si dimostra cieco ai rischi sistemici, come l'impatto ambientale e l'erosione dei processi democratici. In contrapposizione, il saggio propone un cambio di paradigma verso un "*costituzionalismo algoritmico sostanziale*". Questo approccio si fonda su due pilastri: in primo luogo, il riconoscimento di un "*diritto fondamentale all'interazione umana*", quale presidio invalicabile della dignità e dell'uguaglianza sostanziale (artt. 2 e 3 Cost.), che va oltre la mera sorveglianza umana (*human oversight*); in secondo luogo, l'obbligo di integrare i principi costituzionali (dignità umana e uguaglianza sostanziale) direttamente nella progettazione dei sistemi (*by design*), rendendo la compatibilità costituzionale un requisito preventivo e non un controllo *ex post*. Solo un modello di sviluppo tecnologico così radicato nei diritti umani può riaffermare la primazia della persona e realizzare un'autentica sovranità algoritmica europea.

Abstract (En): *this essay critically analyzes the European regulatory framework on Artificial Intelligence, particularly Regulation (EU) 2024/1689 (the AI Act) and the related Italian legislative proposal. It is argued that the legal architecture, founded on a purely functional definition of AI and a risk-based approach, is structurally inadequate to ensure the effective protection of fundamental rights. By overlooking the crucial issue of algorithmic opacity (the black box), this model "de-constitutionalizes" fundamental rights, transforming them from inviolable principles into negotiable variables in a cost-benefit calculation. Furthermore, it proves blind to systemic risks, such as environmental impact and the erosion of democratic processes. In opposition, the essay proposes a paradigm shift towards "substantive algorithmic constitutionalism". This approach rests on two pillars: first, the recognition of a fundamental "right to human interaction" as an inviolable safeguard of human dignity and substantive equality (Arts. 2 and 3 of the Italian Constitution), which transcends mere human oversight; second, the obligation to integrate constitutional principles (human dignity and substantive equality) directly into the design of AI systems (by design), thereby making constitutional compatibility a preventive requirement rather than an* ex post *control. Only such a model of technological development, rooted in human rights, can reaffirm the primacy of the person and achieve genuine European algorithmic sovereignty.*

Parole chiave: *Costituzionalismo algoritmico sostanziale - Diritto all'interazione umana - Approccio basato sul rischio (Critica) - Opacità algoritmica (*Black Box*)* - AI Act.

Keywords: *Substantive Algorithmic Constitutionalism - Right to Human Interaction - Risk-Based Approach (Critique) - Algorithmic Opacity (Black Box) - AI Act.*

1. Introduzione: la tensione tra innovazione tecnologica e garanzie costituzionali.

L'avvento dell'Intelligenza Artificiale (IA) non rappresenta una mera evoluzione tecnologica, ma una trasformazione paradigmatica che interroga le fondamenta stesse del diritto costituzionale[1].

Già la locuzione stessa "*Intelligenza Artificiale*", onnipresente nel dibattito pubblico e nel contesto normativo, cela una mera illusione antropomorfica.

Non siamo però, almeno per adesso, di fronte a una vera e propria nuova forma di intelligenza, ma a sistemi automatizzati, iper-tecnologici, che simulano l'intelligenza umana, perché producono risultati (financo decisioni) "*intelligenti*" - magari più ingegnosi anche perché più rapidi, sintetici, pieni di risorse istantanee di quelli degli esseri umani, a fronte di analoghi *input* - e che sono convenzionalmente, suggestivamente e sinteticamente definiti di "*Intelligenza Artificiale*". E ciò fanno con meccanismi del tutto diversi dalle logiche di pensiero umane e agli esseri umani sempre più impenetrabili e inaccessibili.

La loro caratteristica fondante non è, infatti, il "*pensiero*" in senso umano - inteso come comprensione semantica e coscienza - ma un'autonomia e un'adattabilità crescenti, basate sulla "*manipolazione*" di incontabili dati auto-appresi. Questa stessa capacità (questa sì

---

[1] [*] *Ricercatore e professore aggregato di Diritto costituzionale e pubblico nell'Università degli studi di Catania.*

[**] Questo saggio è l'esito delle ricerche e degli studi condotti nell'ambito del progetto di ricerca Responsabilità, Complessità, Tecnologie (Re.Com.Te.), finanziato dall'Università degli Studi di Catania con il Piano di incentivi per la ricerca di Ateneo 2024/2026 (Pia.ce.ri.) linea 1 (Progetti di ricerca collaborativa) ed è destinato alla prossima pubblicazione nel volume "*Complessità e Scienze sociali*".

Cfr. F. DONATI, *Intelligenza artificiale e giustizia*, in *Rivista AIC*, 1, 2020, 415 ss.; A. SIMONCINI, *Il linguaggio dell'Intelligenza Artificiale e la tutela costituzionale dei diritti*, in *Rivista AIC*, 2, 2023, 1 ss.; O. POLLICINO, *Regolazione e innovazione tecnologica nell'"ordinamento della rete"*, in *Rivista AIC*, 2, 2025, 119 ss.; O. POLLICINO, M. BASSINI e G. DE GREGORIO, *Un «diritto al digitale»?*, in L. Violante e A. Pajno (a cura di), *Biopolitica, pandemia e democrazia*, II, Bologna, Il Mulino, 2021.

"*artificiale*") li rende "*opachi*" per natura. Certo, anche l'intelligenza umana non è un libro aperto: le nostre decisioni sono sovente il frutto di intuizioni, emozioni e processi inconsci che sfuggono a una piena introspezione razionale. Tuttavia, l'opacità bio-fenomenologica dell'essere umano è di natura fondamentalmente diversa dalla opacità puramente computazionale dei sistemi di IA. La loro "*opacità*" algoritmica non è un difetto tecnico, ma una caratteristica strutturale, un nodo gordiano di complessità tale che, pur essendo in teoria tracciabile, è in pratica inintelligibile. È proprio questa differenza qualitativa a scavare l'abisso che separa i loro processi statistici e disincarnati dall'intelligenza umana, che è invece intrinsecamente incarnata, emotiva e radicata nell'esperienza soggettiva. Un abisso destinato ad ampliarsi man mano che la loro complessità esponenzialmente aumenta, senza poter accedere, almeno fino ad oggi, al significato, alla coscienza e alla responsabilità morale che definiscono il giudizio umano.

Comprendere e probabilmente accettare questa natura intrinsecamente e strutturalmente insondabile e sfuggente è il presupposto indispensabile per valutare criticamente ogni tentativo di regolamentazione normativa, a partire da quello europeo. La capacità dei sistemi di IA di operare con un grado sempre più crescente di autonomia, la loro propria opacità tecnica - una caratteristica peculiare dei moderni sistemi di apprendimento automatico, a differenza della tradizionale IA "*simbolica*" basata su regole - ma anche commerciale o potenzialmente *intenzionale* e la loro prevedibile maggiore pervasività in ambiti decisionali pubblicistici (riferimento della nostra indagine) - legislativo[2], amministrativo[3] e giurisdizionale[4] - che possono incidere, e profondamente, sui diritti fondamentali della persona[5], impongono un ripensamento critico dei modelli regolatori finora proposti.

---

[2] Il 9 luglio scorso, alla Camera dei deputati sono stati presentati i primi tre prototipi di Intelligenza Artificiale generativa destinati a supportare i lavori parlamentari, la trasparenza e il rapporto con i cittadini. L'iniziativa mira a integrare l'innovazione tecnologica nei processi democratici con un approccio umano-centrico. L'evento si è aperto con una *lectio magistralis* del professor Luciano Floridi (Università di *Yale*), il quale ha definito la Camera all'avanguardia a livello mondiale, sottolineando l'opportunità per l'Italia di guidare l'Europa in questo settore. I tre prototipi presentati, frutto della collaborazione con diverse università e centri di ricerca italiani, sono: *NORMA*: un assistente virtuale per l'analisi della produzione legislativa. Utilizzando il linguaggio naturale, interroga gli archivi della Camera per fornire dati, anche in formato grafico, sull'*iter* delle leggi e sugli emendamenti. È sviluppato a partire da una proposta del Politecnico di Milano e dell'Istituto Einaudi. *MSE* (Sistema di Scrittura Assistita): uno strumento pensato per supportare deputate e deputati nella redazione di emendamenti, facilitando l'accesso alle fonti interne della Camera e velocizzando la modifica dei testi. Nasce da una proposta del consorzio Alma Human AI (guidata dall'Università di Bologna). *DEPUCHAT*: un *chatbot* rivolto ai cittadini per facilitare l'accesso alle informazioni sull'attività parlamentare dei singoli deputati (proposte di legge, interventi, atti ispettivi). Il sistema, ancora in fase sperimentale, utilizza dati certificati e non risponde a domande sulla sfera personale. È basato su un'idea delle Università di Roma Tre e di Firenze. La Vicepresidente Ascani ha sottolineato come questa iniziativa rappresenti una "*prova di tenuta*" per l'istituzione, con l'obiettivo di governare il cambiamento tecnologico per preservare la centralità dell'essere umano e rafforzare la democrazia. In *https://comunicazione.camera.it/archivio-prima-pagina/19-51129*.

[3] L'algoritmo Ve.R.A. (Verifica Rapporti Finanziari) dell'Agenzia delle Entrate rappresenta il più maturo esempio di IA operativa su larga scala, impiegato con successo nella lotta all'evasione fiscale attraverso l'analisi predittiva del rischio. Parallelamente, l'INPS si distingue per l'uso di assistenti virtuali come DOT e per la sperimentazione di sistemi di IA a supporto delle politiche attive del lavoro.

[4] La frontiera della giustizia predittiva è esplorata attraverso progetti pilota, come quello della Corte d'Appello di Venezia e il progetto "*IustitIA*" della Corte d'Appello di Reggio Calabria, che mirano ad analizzare la giurisprudenza per aumentare la prevedibilità delle decisioni.

[5] F. DONATI, *La protezione dei diritti fondamentali nel Regolamento EU sull'intelligenza artificiale*, in *Rivista AIC*, 1, 2025, 2; C. NAPOLI, *Algoritmi, intelligenza artificiale e formazione della volontà pubblica: la decisione*

Questa sfida si inserisce direttamente nel solco del costituzionalismo contemporaneo, la cui funzione essenziale è sempre stata quella di limitare il potere per tutelare i diritti. Se il costituzionalismo del XX secolo è nato per porre un limite giuridico al potere politico dello Stato e alla "*tirannia della maggioranza*", quale nuovo assolutismo, attraverso costituzioni rigide e il controllo di costituzionalità[6], oggi la sfida è adattare quegli stessi principi per porre un limite a una nuova forma di potere, diffuso e opaco: il potere algoritmico[7].

La questione non è dunque e ad una prima sommaria riflessione creare un diritto nuovo dal nulla, ma estendere la portata e la sostanza delle garanzie costituzionali esistenti alla dimensione tecnologica, e, più a monte, non è più *se* regolare l'IA, ma *come* farlo in modo che la regolazione stessa non si traduca in una abdicazione ai principi cardine dello Stato costituzionale[8] di diritto[9].

Questo saggio si propone di verificare criticamente l'adeguatezza del quadro normativo emergente, rappresentato a livello sovranazionale dal Regolamento (UE) 2024/1689 (*Artificial Intelligence* o *AI Act*)[10] e, nell'ordinamento interno e in prospettiva prossima, dal disegno di legge, attualmente all'esame del Senato, n. 1146-B[11].

La comune architettura giuridica, fondata su una definizione legale puramente funzionale dell'IA e su un approccio basato sulla gestione del rischio (*risk-based approach*), sembra possa rivelarsi strutturalmente insufficiente a garantire una tutela sostanziale ed effettiva dei diritti umani.

---

amministrativa e quella giudiziaria, in *Rivista AIC*, 3, 2020, 319.

[6] R. BIN e G. PITRUZZELLA, *Diritto costituzionale*, XXV, Giappichelli, Torino 2024, 13.

[7] Questo potere non è meramente tecnico, ma si inscrive in quella che Julie E. Cohen ha definito la progressiva costruzione di un "*capitalismo informazionale*", in cui le piattaforme digitali non si limitano a processare dati, ma plasmano attivamente le soggettività e le relazioni sociali attraverso le loro architetture di sorveglianza e modulazione comportamentale. La regolazione dell'IA, in questa prospettiva, non riguarda solo la gestione di una tecnologia, ma la necessità di costituzionalizzare un potere privato che opera su scala globale. Cfr. J. E. COHEN, *Configuring the Networked Self: Law, Code, and the Play of Everyday Practice*, Yale University Press, 2012, e, più recentemente, *Between Truth and Power: The Legal Constructions of Informational Capitalism*, Oxford University Press, 2019.

[8] A. BARBERA, C. FUSARO e C. CARUSO, *Corso di diritto costituzionale*, VII, Il Mulino, Bologna, 2024, 46 ss., sottolineano come l'espressione stato costituzionale indichi una forma di stato che nella costituzione trova la propria stessa identità.

[9] Per una ricostruzione diacronica dell'evoluzione della regolazione europea, dal "*liberismo digitale*" della Direttiva 2000/31/CE sul commercio elettronico a una nuova stagione di "*costituzionalismo digitale*" segnata da interventi come il GDPR, *General Data Protection Regulation*, Regolamento (UE) 2016/679, Regolamento generale sulla protezione dei dati, il DSA, *Digital Service Act*, Regolamento (UE) 2022/2065, e l'*AI Act*, di cui alla nota seguente, si veda O. POLLICINO, *Regolazione e innovazione tecnologica*, cit., 137 ss.

[10] Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale, in GU L 2024/1689 del 12 luglio 2024, già applicabile dal 2 febbraio 2025 e dal 2 agosto 2025 in parte, ex art. 113.

[11] D.d.l. A.S. n. 1146-B, *Disposizioni e deleghe al Governo in materia di intelligenza artificiale*, presentato dal Presidente del Consiglio dei Ministri (Meloni) e dal Ministro della Giustizia (Nordio), approvato dal Senato della Repubblica il 20 marzo 2025 e approvato con modificazioni dalla Camera dei deputati il 25 giugno 2025, attualmente all'esame del Senato. In *https://www.senato.it/leggi-e-documenti/disegni-di-legge/scheda-ddl?did=59313*. Cfr. O. POLLICINO, *Regolazione e innovazione tecnologica*, cit., 120 ss., il quale descrive la transizione da una fase di "*liberismo digitale*" a una di "*costituzionalismo digitale*". In quest'ottica, l'*AI Act*, pur essendo l'ultimo stadio di tale evoluzione, di fronte a una sfida tecnologica più complessa (l'autonomia dell'IA) rischia di configurare una reazione legislativa meno garantista rispetto a quella che ha caratterizzato la precedente "*stagione dell'automazione algoritmica*" (es. il DSA).

Tale architettura normativa sposa, a monte, una scelta definitoria che ignora il nucleo problematico della tecnologia che intende regolare, inevitabilmente condizionandone, a valle, l'intero impianto regolatorio. Essa nasce da quello che si potrebbe definire un "*peccato originale*": una scelta definitoria che, sottovalutando il centro più complesso della tecnologia artificiale contemporanea, ne condiziona inevitabilmente l'intero impianto regolatorio.

La logica del rischio, infatti, tende a trasformare i diritti fondamentali da principi assiologici inviolabili a variabili ponderabili in un prevalente calcolo di convenienza, normalizzando la possibilità della loro compressione in nome dell'innovazione. Questo cedimento non è una mera scelta tecnica, ma una significativa decisione politica che rischia di regolarizzare uno stato di "*eccezione permanente*", in cui i diritti non sono più il limite invalicabile del potere, ma un costo gestibile del progresso tecnologico.

Una conferma quasi empirica di questa deriva "*de-costituzionalizzante*" (termine qui usato provocatoriamente e suggestivamente nel senso quantomeno di un depotenziamento delle garanzie dei diritti umani) emerge dagli stessi recenti primi orientamenti (*Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689*)[12] che la Commissione Europea ha adottato per chiarire l'applicazione delle pratiche vietate. Nel dettagliare il divieto di manipolazione e sfruttamento dannosi (art. 5, par. 1, lett. a e b, *AI Act*), gli orientamenti specificano che una pratica è illecita solo se provoca o può ragionevolmente provocare un "*danno significativo*" (fisico, psicologico, finanziario ed economico).

Lungi dall'essere un mero dettaglio tecnico, questa soglia rappresenta il cuore della logica basata sul rischio: non si vieta la manipolazione in sé, in quanto intrinsecamente lesiva della dignità e dell'autonomia individuale, ma solo quella che supera un determinato livello di gravità, valutato "*caso per caso*" in base a criteri come la portata, l'intensità e la durata del danno. In questo modo, gli orientamenti stessi ammettono l'esistenza di un'area di manipolazione e sfruttamento "*leciti*" - quelli che non raggiungono la soglia del "*danno significativo*" - trasformando di fatto principi costituzionali inviolabili in costi gestibili dell'innovazione, esattamente come paventato.

Come ammoniva Stefano Rodotà, in un'epoca di trasformazioni così radicali, la sfida è quella di preservare "*il diritto di avere diritti*", ovvero la precondizione stessa della dignità e della cittadinanza[13]. In contrapposizione a tale modello, potrebbe essere opportuno, se non necessario, un cambio di paradigma verso una regolazione di stampo sostanziale, preventiva e radicalmente antropocentrica. Un approccio che non si limiti alla gestione dei possibili "*danni*" *ex post*, ma che possa governare la tecnologia *ex ante*, ancorandola maggiormente ed effettivamente ai principi supremi dell'ordinamento costituzionale, quali, innanzi tutto, la dignità della persona, sancita dall'art. 2 Cost., e l'uguaglianza sostanziale, come delineata dall'art. 3, comma 2, Cost. Si tenterà, quindi, di valutare criticamente la definizione funzionale di IA adottata dal legislatore europeo, a cui pare voglia rinviare pedissequamente quello italiano, che costituisce il fondamento logico dell'approccio basato sul rischio,

[12] Si tratta degli Orientamenti della Commissione relativi alle pratiche di intelligenza artificiale vietate ai sensi del regolamento (UE) 2024/1689 (regolamento sull'IA), pubblicati il 4 febbraio 2025 in *https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act*.

[13] S. RODOTÀ, *Il diritto di avere diritti*, Bari, Laterza, 2012. L'opera esplora l'evoluzione della dignità umana e la necessità di un diritto rinnovato per affrontare le sfide della globalizzazione e delle nuove tecnologie.

tentando di evidenziarne la difficoltà di cogliere a pieno il nucleo problematico dell'opacità algoritmica; e lo stesso approccio euro-unionale basato sul rischio, vagliandone l'adeguatezza a fronteggiare non solo i rischi individuali, ma soprattutto quelli di natura sistemica - come il prevedibile impatto ambientale e la potenziale erosione dei processi democratici - che vengono quasi interamente trascurati; e la sua tendenza a "*de-costituzionalizzare*" diritti fondamentali, che possono così essere solo salvaguardati *ex post*, solo con la tecnica di tutela dei contro limiti e quindi per il tramite del controllo di legittimità della Corte costituzionale o, ancor prima, tramite interpretazioni giurisprudenziali conformi alla Costituzione, caso per caso, e dunque frammentariamente.

Si proverà a delineare i fondamenti di un approccio alternativo ad altri possibili e cumulabili paradigmi regolatori, basato sul riconoscimento di un fondamentale "*diritto all'interazione umana*", autorevolmente enucleato e sostenuto come presidio della dignità umana e dell'uguaglianza sostanziale[14].

Infine, si esamineranno criticamente, seppur sommariamente, anche le garanzie procedurali e giurisdizionali, i possibili limiti attuali e si abbraccerà un modello di "*costituzionalismo algoritmico sostanziale*"[15], come via per assicurare una tutela effettiva dei diritti inviolabili dell'essere umano.

Le conclusioni intendono tracciare le linee per un modello di sviluppo tecnologico radicato *nei* diritti umani, unica via, ci sembra, per una vera "*sovranità algoritmica*" europea, quale capacità dell'Unione Europea di garantire che lo sviluppo, l'utilizzo e la *governance* dei sistemi algoritmici e dell'intelligenza artificiale sul suo territorio siano allineati ai propri valori fondamentali, ai diritti dei cittadini e ai suoi interessi strategici e, in definitiva, volontà di plasmare un futuro digitale che rifletta i valori europei, garantendo che la tecnologia rimanga uno strumento al servizio dell'uomo e della società, e non il contrario.

2. La definizione legale dell'IA tra funzionalismo e opacità: un fondamento insufficiente per la tutela dei diritti fondamentali?

Il punto di partenza di ogni tentativo di regolazione è la definizione del suo oggetto.

Sia consentito un breve raffronto testuale tra l'*AI Act* e il disegno di legge italiano, certamente meritevole di approfondimento in altre sedi, se e quando quest'ultimo diventerà legge.

L'art. 3, punto 1), dell'*AI Act* definisce "*«sistema di IA»: un sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, deduce dall'input che riceve come generare output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali*".

[14] Cfr. G. SCACCIA e A. MONORITI, *Quali spazi per un diritto all'interazione umana?*, in *Federalismi.it*, Editoriale, 15 gennaio 2025.

[15] Che converge con il paradigma dell'"*Algorithm Constitutional by design*", cfr. G. DE MINICO, *Towards an "Algorithm Constitutional by Design"*, in *BioLaw Journal* - Rivista di BioDiritto, 1, 2021, 381 ss., ove in conclusione si legge "*una regolamentazione vincolante, seppur ridotta al minimo, potrà disegnare un algoritmo conforme ai valori costituzionali europei, in altri termini un "algoritmo costituzionale per progettazione". In una prospettiva più generale, orienterà la tecnologia verso un bene comune equo e diffuso nel rispetto di un quadro istituzionale democratico*". L'approccio "*by design*" è stato reso celebre da Ann Cavoukian, che negli anni Novanta ha sviluppato il concetto di "*Privacy by Design*", un principio secondo cui la protezione dei dati personali deve essere integrata nella progettazione di sistemi e tecnologie fin dall'inizio, e non aggiunta come un elemento accessorio. Cfr. anche M. HILDEBRANDT, *Legal Protection by Design. Objections and Refutations*, in *Legisprudence*, 2, 2011, 223 ss.

Il disegno di legge italiano, all'art. 2, comma 1, lett. a), opera un rinvio recettizio a tale nozione.

La relazione tra i due testi è di piena identità e di totale incorporazione della definizione europea. In più, l'art. 2, comma 2, rimanda, per quanto non espressamente previsto, alle altre definizioni del regolamento ed è anch'essa di tendenziale completo allineamento, con prevenzione di qualsiasi disarmonia terminologica. Già la scelta definitoria, volutamente ampia e tecnologicamente neutra per resistere nel tempo, appare problematica sotto il profilo costituzionale.

Essa si concentra sulle *capacità* tecniche della macchina - autonomia, adattabilità, inferenza - trascurando la sua *funzione* giuridicamente (e socialmente) più dirompente: la capacità di assumere, direttamente o indirettamente, decisioni in modo sempre più autonomo, ma sempre più opaco, ciò che può incidere sulla sfera giuridica dei singoli e della collettività.

Come acutamente osservato, la peculiarità dell'IA non risiede nell'eseguire compiti, ma nella possibilità sempre più concreta e attuale di decidere, un'attività finora riservata all'essere umano. Questo scarto paradigmatico mette in crisi le categorie giuridiche tradizionali di "*mezzo-fine*" e di "*agente-strumento*", su cui si fondano istituti cardine come la responsabilità e l'imputabilità[16].

La dottrina ha già messo in luce come una definizione così generica possa generare significative incertezze applicative, creando zone d'ombra regolatorie[17]. Il vizio più profondo di questa impostazione funzionalista, tuttavia, sembra risiedere nell'ignorare la caratteristica strutturale che distingue i moderni sistemi di IA: l'opacità del processo decisionale[18].

Questa impostazione ignora una distinzione fondamentale all'interno del campo dell'IA, che è la vera radice del problema. Storicamente, il paradigma dominante era quello dell'IA "*simbolica*". In questo modello, la conoscenza umana veniva esplicitamente codificata sotto forma di regole logiche e simboli. Il sistema non apprendeva dai dati, ma operava seguendo un percorso deduttivo basato su istruzioni pre-programmate (ad esempio, i sistemi esperti basati su una logica "*se-allora*"). La caratteristica fondamentale di questo approccio era la sua intrinseca trasparenza: il suo processo decisionale, essendo il risultato dell'applicazione di regole definite dall'uomo, era interamente tracciabile e spiegabile.

La transizione da un'IA "*simbolica*", basata su regole logico-deduttive preimpostate dall'uomo (logica della causazione), a un'IA "*sub-simbolica*", fondata sull'apprendimento automatico (*machine learning*) e su quello profondo (*deep learning*), ha generato sistemi che

[16] A. SIMONCINI, *Il linguaggio dell'Intelligenza Artificiale*, cit., 8 ss., il quale descrive come la capacità dell'IA di "*prendere*" decisioni, e non solo di "*eseguirle*", metta in crisi le tradizionali categorie giuridiche fondate sulla distinzione tra "*agente-strumento*" e "*mezzo-fine*".

[17] O. POLLICINO, *Regolazione e innovazione tecnologica*, cit., 159-160, che evidenzia come ambiguità definitorie possano condurre a fenomeni di *sovra-regolazione* di ostacolo all'innovazione o di *sotto-regolazione* che favoriscono strategie elusive, quali l'auto-classificazione da parte di aziende di propri prodotti come *software* tradizionali per evitare gli oneri normativi.

[18] Cfr. D. U. GALETTA e J. G. CORVALÁN, *Intelligenza Artificiale per una Pubblica Amministrazione 4.0? Potenzialità, rischi e sfide della rivoluzione tecnologica in atto*, in *federalismi.it*, 3, 2019, 15, i quali evidenziano il fenomeno della "*scatola nera*" (*black box*) come il vero problema connesso all'implementazione di sistemi di IA, per cui "*si ritiene che sia praticamente impossibile stabilire in che modo l'Algoritmo di Machine Learning sia giunto ad un certo risultato*". Si veda anche D. U. GALETTA, *Human-stupidity-in-the-loop? Riflessioni (di un giurista) sulle potenzialità e i rischi dell'Intelligenza Artificiale*, in federalismi.it, 5, 2023, VII, la quale spiega che gli algoritmi di *machine learning* giungono a conclusioni "*sulla base di rappresentazioni probabilistiche e di metodi di apprendimento statistici*", rendendo l'esplicabilità *ex post* intrinsecamente problematica.

apprendono da enormi quantità di dati, identificando correlazioni statistiche che sfuggono alla comprensione umana[19].

L'*output* di una rete neurale artificiale, ad esempio - progettata per simulare il funzionamento del cervello umano, utilizzando nodi interconnessi che elaborano le informazioni in modo gerarchico, e capace di apprendere da grandi quantità di dati e di risolvere problemi complessi in tempi rapidissimi - non è il risultato di un'applicazione sillogistica di regole, ma l'esito deterministico di un modello i cui parametri sono il prodotto di un processo di addestramento stocastico; tale processo, pur essendo governato da leggi probabilistiche, genera una catena causale di una complessità tale da renderla, in pratica, non spiegabile in termini causali-deterministici e non ricostruibile in termini umanamente intelligibili[20].

Questa opacità, la cosiddetta *black box*, non è un difetto tecnico emendabile, ma una caratteristica intrinseca di tali sistemi, come magistralmente spiegato e argomentato, e già da tempo, da uno dei maggiori studiosi di diritto dell'intelligenza artificiale[21], che mina alla base i principi di trasparenza, di motivazione del provvedimento e, in ultima istanza, di giustiziabilità della decisione[22].

---

[19] Sul punto, si veda sempre A. SIMONCINI, *Il linguaggio dell'Intelligenza Artificiale*, cit., 10 ss. che individua la "*nuova primavera*" dell'IA nel passaggio da un approccio logico-deduttivo, basato sulla "*causazione*", a un modello statistico-induttivo fondato sulla "*correlazione*", il quale, pur garantendo previsioni accurate, rende il processo decisionale intrinsecamente opaco e non spiegabile in termini causali.

[20] Cfr. G. F. LICATA, *Intelligenza artificiale e contratti pubblici: problemi e prospettive*, in *CERIDAP*, 2, 2024, 34; F. DONATI, *Intelligenza artificiale e giustizia*, cit., 416; G. DE MINICO, *Giustizia e intelligenza artificiale: un equilibrio mutevole*, in *Rivista AIC*, 2, 2024, 86.

[21] Come ha magistralmente osservato F. PASQUALE, *The Black Box Society: The Secret Algorithms that Control Money and Information*, Cambridge-London, Harvard University Press, 2015, l'opacità non è un semplice attributo tecnico, ma una strategia di potere. Le potenti organizzazioni di Silicon Valley e Wall Street abusano del segreto - sia esso legale, come il segreto commerciale, o di fatto, dovuto alla complessità del codice - per massimizzare i profitti e sottrarsi alla responsabilità. In questa "società della scatola nera", i cittadini sono resi trasparenti e soggetti a un monitoraggio pervasivo, mentre i meccanismi algoritmici che prendono decisioni cruciali sulle loro vite - dall'accesso al credito alla visibilità delle informazioni - rimangono impenetrabili. L'opacità, quindi, non è un effetto collaterale, ma il nucleo del problema giuridico e politico posto dall'IA. Frank Pasquale è professore di diritto presso la *Cornell Tech* e la *Cornell Law School*. È un esperto di diritto dell'intelligenza artificiale (IA), algoritmi e apprendimento automatico. Tra i suoi libri figurano anche *New Laws of Robotics* (Harvard University Press, 2020). Ha pubblicato oltre 70 articoli su riviste e capitoli di libri, e ha co-curato *The Oxford Handbook on the Ethics of Artificial Intelligence* (Oxford University Press, 2020) e *Transparent Data Mining for Big and Small Data* (Springer-Verlag, 2017).

[22] Il riferimento è sempre a Frank Pasquale, v. nota precedente. Pasquale ha analizzato come potenti attori privati utilizzino algoritmi segreti per prendere decisioni cruciali in settori come la finanza, il marketing e la reputazione *online*, creando una "*società della scatola nera*" in cui le logiche del potere sono inaccessibili e incontestabili. Mentre l'analisi di Pasquale si concentra sulla necessità di una maggiore *accountability* e trasparenza per contrastare questo potere privato, il presente saggio tenta di svilupparne le implicazioni sul piano del diritto costituzionale, sostenendo che l'opacità non è solo un problema di *governance*, ma una minaccia diretta alla sostanza stessa dei diritti fondamentali e alla tenuta dello Stato di diritto. F. DONATI, *Intelligenza Artificiale e Giustizia*, cit., 421 ss., il quale analizza approfonditamente il caso del *software* COMPAS e i rischi di discriminazioni algoritmiche derivanti sia da errori nei dati di addestramento sia dalla "*discriminazione statistica*". La critica all'opacità come caratteristica intrinseca dei moderni sistemi di IA, che mina i principi di trasparenza e giustiziabilità, è un tema centrale nella dottrina costituzionalistica. C. CASONATO, *Costituzione e intelligenza artificiale: un'agenda per il prossimo futuro*, in *Consulta Online*, 13 gennaio 2020, 7, ha evidenziato come il fenomeno della *black box* - l'impossibilità di comprendere il percorso logico che conduce a un *output* - sia particolarmente problematico in settori come la giustizia, dove rende ogni decisione automatizzata "*attaccabile in quanto non provvista della necessaria motivazione (art. 111 Cost.)*". Anche T. GROPPI, *Alle frontiere dello Stato Costituzionale: innovazione tecnologica e intelligenza artificiale*, in *Consulta Online*, III,

L'intero impianto regolatorio dell'*AI Act* nasce, pertanto, con un vizio d'origine. Fondandosi su una definizione che astrae dalla questione cruciale dell'opacità, esso è logicamente costretto a un approccio "*correttivo*" e successivo.

Le stesse linee guida sull'ambito di applicazione degli obblighi per i modelli di IA per scopi generali della Commissione Europea[23], pur tentando di fornire criteri pratici per identificare i modelli di IA per finalità generali (*General-Purpose AI* - GPAI), ricadono in questa logica funzionalista. Essa propone un criterio indicativo basato sulla potenza di calcolo impiegata per l'addestramento (maggiore di 10^{23} FLOP) e sulla capacità di generare linguaggio, immagini o video. Questo approccio, sebbene pragmatico, si concentra sulla "*capacità*" e sulla "*scala*" del modello, tralasciando la questione fondamentale di come tale capacità si genera e si manifesta, ovvero il problema dell'opacità intrinseca.

La definizione funzionale sembra rassegnarsi all'opacità come caratteristica intrinseca e non governabile della tecnologia e accettare la *black box* come un dato di fatto e conseguentemente l'unica opzione regolatoria risulta essere stata quella di spostare l'attenzione dall'*interno* (dalla logica del sistema) all'*esterno* (agli effetti del sistema), dalla *causa* (come funziona l'algoritmo) al *danno* (quali rischi produce il suo *output*).

Anziché imporre requisiti strutturali *ex ante* sulla progettazione stessa dei sistemi, pretendendo una tecnologia intrinsecamente comprensibile, il legislatore si trova a voler

---

2020, 679, sottolinea come i nuovi sistemi operino "*nell'oscurità e nel segreto degli algoritmi*", in contrasto con il principio democratico del potere visibile. La nostra critica alla definizione funzionale di IA adottata dall'*AI Act*, che ignora il nodo cruciale dell'opacità, trova un solido alleato nell'analisi di G. FASANO, *Le 'informazioni sintetizzate' generate dai large language models e le esigenze di tutela del diritto all'informazione: valori costituzionali e nuove regole*, in *dirittifondamentali.it*, 1, 2024, 107-132. L'autore chiarisce che, sebbene l'*output* dei *large language models* non sia qualificabile come "*manifestazione di pensiero*", esso produce e veicola "*informazioni*" che incidono profondamente sulla libertà di formazione del convincimento interiore degli individui. Questa riflessione è fondamentale, poiché sposta il focus dal "*come*" l'algoritmo funziona al suo impatto sulla persona, in linea con il nostro approccio sostanziale. Inoltre, la critica dell'autore all'idea di una vera "*autonomia*" della macchina, da lui ricondotta a una sofisticata eteronomia che non esclude la responsabilità umana, supporta direttamente la nostra tesi contro l'accettazione della *black box* come un dato di fatto ineluttabile. Per rispondere al problema dell'opacità è sorto, da anni, un intero settore di ricerca noto come Intelligenza Artificiale Spiegabile (XAI). L'obiettivo della XAI è sviluppare tecniche per rendere comprensibili le decisioni dei modelli opachi. Le metodologie più note, come LIME (*Local Interpretable Model-agnostic Explanations*) e SHAP (*SHapley Additive exPlanations*), operano *ex post*, ovvero dopo che il modello è stato addestrato. Esse tentano di spiegare una singola previsione creando un modello locale, più semplice e interpretabile (es. una regressione lineare), che "*riduce*" il comportamento della *black box* solo intorno di quella specifica decisione. Tuttavia, tali approcci presentano limiti noti: le spiegazioni fornite possono essere instabili (cambiando anche a fronte di piccole perturbazioni dell'*input*) e, soprattutto, non vi è garanzia che siano pienamente "*fedeli*" alla reale logica interna del modello complesso che cercano di emulare. Questa intrinseca debolezza delle soluzioni puramente correttive e *a posteriori* offre un'ulteriore conferma della necessità, sostenuta nel presente saggio, di un approccio preventivo e strutturale, che imponga la trasparenza *by design*. Il testo che ha introdotto e formalizzato il concetto di spiegazione locale agnostica rispetto al modello, diventando un punto di riferimento per le tecniche successive come SHAP, è l'articolo originale su LIME di M. T. RIBEIRO, S. SINGH, e C. GUESTRIN, *"Why Should I Trust You?": Explaining the Predictions of Any Classifier*, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, 2016, 1135-1144. Un testo fondamentale che critica esplicitamente l'affidamento a spiegazioni *ex post* per decisioni critiche e che argomenta con forza a favore di modelli intrinsecamente interpretabili è quello di C. RUDIN, *Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead*, in *Nature Machine Intelligence*, 1, *may* 2019, 206-215.

[23] Si tratta delle *Guidelines on the scope of the obligations for general-purpose AI models established by Regulation (EU) 2024/1689*, pubblicate il 18 luglio 2025, insieme al *Code of Practice for General-Purpose AI Models*, in *https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai*.

risolvere i problemi - i rischi, le discriminazioni, le violazioni dei diritti - *ex post* con strumenti correttivi, che, come si vedrà, possono rivelarsi intrinsecamente deboli.

L'approccio basato sul rischio non sembra essere, dunque, una scelta adeguatamente ponderata tra diverse opzioni regolatorie, ma la conseguenza inevitabile di una definizione iniziale probabilmente manchevole, che rinuncia a governare la natura della tecnologia per limitarsi a gestirne gli effetti.

Se il legislatore avesse posto al centro della definizione legale il problema dell'opacità, riconoscendola come la principale minaccia alla trasparenza, alla giustiziabilità e alla responsabilità, l'architettura regolatoria di settori in cui tali principi sono giuridicamente rilevanti avrebbe potuto (e dovuto) essere radicalmente diversa. Una definizione incentrata sull'opacità avrebbe imposto, per coerenza logica, requisiti di "*spiegabilità per progettazione*" (*explainability by design*)[24]. In un tale scenario, il diritto non si adatterebbe alla tecnica, ma la tecnica dovrebbe essere progettata per essere costituzionalmente compatibile[25]. L'approccio basato sul rischio, se non del tutto superfluo, sarebbe diventato secondario: un meccanismo residuale per gestire i rischi non eliminabili attraverso la progettazione.

La scelta di una definizione funzionale e tecnologicamente neutra, quindi, sembra essere un atto di accettazione delle condizioni poste dalla tecnologia e dai suoi sviluppatori, con rinuncia a plasmarne la natura in conformità con i principi fondamentali dello Stato costituzionale di diritto. Ciò che pare condizionare l'intera logica dell'*AI Act*, costringendolo a una rincorsa perpetua e intrinsecamente debole contro i danni, anziché (tentare di) prevenirli o ridurli alla fonte[26].

Anziché imporre requisiti strutturali di trasparenza e di non manipolabilità *by design*, ci si trova costretti a inseguire gli esiti dannosi di una *black box* che si accetta come un dato di fatto, confermandosi la necessità di un cambio di paradigma verso un costituzionalismo algoritmico che ponga i diritti come vincolo di progettazione e non come danno da mitigare *a posteriori*.

---

[24] Il concetto di "*spiegabilità per progettazione*" (*explainability by design*), qui invocato sul piano giuridico-costituzionale, trova un diretto e fondamentale parallelo nel dibattito tecnico-informatico. Esso si contrappone nettamente ai citati metodi di spiegazione *ex post* (come LIME e SHAP), che tentano di interpretare un modello opaco "*dall'esterno*", e promuove invece lo sviluppo di modelli "*intrinsecamente interpretabili*" (*intrinsically interpretable models*). Questi non si limitano ai sistemi tradizionali (es. alberi decisionali, regressione logistica), le cui decisioni sono trasparenti per costruzione, ma includono una frontiera avanzata della ricerca che mira a creare architetture di *deep learning* che, pur mantenendo un'elevata capacità predittiva, siano progettate per essere trasparenti e comprensibili. L'approccio qui sposato del "*costituzionalismo algoritmico sostanziale*" si allinea quindi con la corrente più esigente della stessa ricerca in XAI, che riconosce i limiti di un approccio puramente correttivo e la necessità di integrare l'interpretabilità nell'architettura stessa dei sistemi. V. M. T. RIBEIRO, S. SINGH, e C. GUESTRIN, *"Why Should I Trust You?"*, cit., 1135-1144, e C. RUDIN, *Stop Explaining Black Box Machine Learning Models*, cit., 206-215.

[25] Questa prospettiva si pone nel solco della celebre tesi di Lawrence Lessig, "*Code is Law*", secondo cui l'architettura tecnologica (il "*codice*") è una forma di regolazione potente, forse più efficace della legge stessa, in quanto plasma le possibilità di azione degli individui nel cyberspazio. Cfr. L. LESSIG, *Code and Other Laws of Cyberspace*, Basic Books, 1999. Se per Lessig il codice può essere uno strumento per implementare o, al contrario, eludere i valori legali, la proposta di un "*costituzionalismo algoritmico sostanziale*" qui avanzata ne radicalizza l'intuizione: non si tratta solo di rendere il codice conforme alla legge, ma di esigere che i principi costituzionali diventino la grammatica fondamentale del codice stesso, un passaggio da una compatibilità esterna a una coerenza interna.

[26] Un rischio ulteriore è che la tecnica stessa, specie se opaca e autogenerativa, si trasformi essa stessa in fonte del diritto (*code as source of law*), dettando regole concrete e *ad personam* senza mediazione umana e al di fuori di ogni processo democratico. Ciò segnerebbe il passaggio dalla *rule of law* a una problematica *rule of techs*, dove la decisione politica è di fatto assorbita dalla determinazione meccanica. Su questa deriva, si veda G. DE MINICO, *Le fonti del diritto: un argine all'intelligenza artificiale?*, in *Rivista AIC*, 3, 2025, 91, 96.

3. Critica all'approccio basato sul rischio.

L'architettura dell'*AI Act* si fonda su una classificazione dei sistemi di IA in base al livello di rischio che essi comportano per la salute, la sicurezza e i diritti fondamentali.

Questa "*piramide del rischio*" articola la disciplina su quattro livelli: rischio inaccettabile (pratiche "*vietate*"[27]), rischio alto (sistemi soggetti a requisiti stringenti), rischio limitato (obblighi di trasparenza) e rischio minimo (nessun obbligo specifico)[28]. Una disciplina a parte e trasversale hanno i modelli di IA per finalità generali (GPAI)[29] che in presenza di determinate condizioni[30], sono considerati a rischio sistemico[31].

Il disegno di legge italiano stabilendo, all'art. 1, comma 2, che la legge si interpreta e si applica "*conformemente al regolamento (UE) 2024/1689*", e specificando, all'art. 3, comma 5, che la legge "*non produce nuovi obblighi rispetto a quelli previsti dal regolamento (UE) 2024/1689 per i sistemi di intelligenza artificiale*", adotta l'intera architettura basata sul rischio del Regolamento UE, con la stessa classificazione e i conseguenti obblighi di conformità. Sebbene l'approccio sul rischio appaia, a prima vista, ispirato a un principio di razionalità e di proporzionalità, un'analisi più approfondita dell'*AI Act* ne rivela la natura formalistica e la sostanziale inadeguatezza a fornire una tutela forte[32].

Il considerando 26 dell'*AI Act* giustifica tale scelta in nome della necessità di introdurre "*un insieme proporzionato ed efficace di regole vincolanti*", adattando "*la tipologia e il contenuto di dette regole all'intensità e alla portata dei rischi che possono essere generati dai sistemi di IA*".

Il principio di proporzionalità, nel diritto europeo, richiede un bilanciamento tra l'obiettivo perseguito e la compressione dei diritti. L'approccio basato sul rischio, tuttavia, opera una profonda e pericolosa mutazione concettuale: "*de-costituzionalizza*" i diritti fondamentali.

I diritti, che nell'architettura costituzionale rappresentano principi assoluti e inviolabili posti a limite invalicabile del potere, vengono ridotti a meri "*rischi*" da "*valutare*", da "*mitigare*" e, implicitamente, da "*accettare*" in un calcolo di bilanciamento con le esigenze dell'innovazione e del mercato[33].

La domanda fondamentale non è più "*questo sistema viola la dignità umana?*", ma diventa: "*qual è il livello di rischio accettabile di violazione della dignità umana?*". Questa inversione di prospettiva tradisce il cuore del costituzionalismo contemporaneo, che si fonda sull'idea di un nucleo di diritti fondamentali "*inviolabili*" (come recita l'art. 2 della Costituzione) che costituiscono un limite invalicabile e non negoziabile per qualsiasi potere, incluso quello legislativo.

---

[27] Art. 5 *AI Act*.

[28] Per un'analisi dettagliata dell'approccio basato sul rischio si vedano gli artt. 6 ss. *AI Act*.

[29] Art. 3, par. 63, *AI Act*.

[30] Art. 51 *AI Act*.

[31] La critica all'approccio basato sul rischio è ampiamente argomentata anche da O. POLLICINO, *Regolazione e innovazione tecnologica*, cit., 160, il quale ne evidenzia il carattere "*top-down*", in contrasto con l'approccio "*bottom-up*" del GDPR, e sottolinea come tale modello, nato per garantire la sicurezza del prodotto nel mercato unico, si riveli un compromesso fragile tra esigenze di mercato e tutela dello Stato di diritto.

[32] O. POLLICINO, *Regolazione e innovazione tecnologica*, cit., 123 ss.; F. DONATI, *La protezione dei diritti fondamentali*, cit., 7 ss.

[33] In senso analogo, E. CIRONE, *L'AI Act e l'obiettivo (mancato?) di promuovere uno standard globale per la tutela dei diritti fondamentali*, in *Quaderni AISDUE*, 2, 2024, 12, 18, la quale osserva come l'*AI Ac*t "*non sembra che rifletta le caratteristiche essenziali per essere considerato uno* standard *globale, proprio in virtù del (solo formalmente centrale) ruolo della protezione dei diritti fondamentali nell'intera struttura del regolamento*". L'autrice critica l'impostazione del regolamento, ispirata alla sicurezza dei prodotti, dove i diritti fondamentali sono indicati come "*interessi pubblici da proteggere assieme alla salute e alla sicurezza*", un approccio che rischia di non assicurarne un'adeguata protezione.

La funzione della Costituzione non è solo quella di "*bilanciare*" i diritti con qualsiasi altro interesse, ma di sottrarne il nucleo a tale bilanciamento, specialmente quando l'interesse contrapposto è puramente di natura economica (e tecnologica).

L'approccio basato sul rischio, al contrario, ripropone una logica utilitaristica che il costituzionalismo del secondo dopoguerra aveva inteso superare, riaffermando la primazia assiologica della persona e della sua dignità. Un approccio che trasforma un diritto inviolabile in una variabile negoziabile non può essere, a nostro avviso, considerato "*proporzionato*" in senso euro-unionale e costituzionale[34]. Esso rappresenta, piuttosto, una resa a un paradigma manageriale-tecnocratico che normalizza la potenziale violazione dei diritti come un costo collaterale e gestibile del progresso tecnologico[35]. Come evidenziato da autorevole dottrina, la rigidità delle categorie di rischio e la problematica delega ad organismi di standardizzazione tecnica a prevalente composizione privata[36], a cui il Regolamento UE affida il compito di sviluppare le norme tecniche armonizzate per supportare l'attuazione dei requisiti previsti, in particolare quelli per i sistemi ad alto rischio, aggravano questo problema, affidando di fatto la tutela dei diritti a soggetti portatori di interessi economici, con evidenti conflitti[37].

---

[34] L'approccio basato sul rischio, cardine dell'*AI Act*, è oggetto di un vivace dibattito. Mentre il presente saggio ne evidenzia la tendenza a "*de-costituzionalizzare*" i diritti, trasformandoli in variabili negoziabili, altri autori offrono una prospettiva differente. E. C. RAFFIOTTA e M. BARONI, *Intelligenza artificiale, strumenti di identificazione e tutela dell'identità*, in *BioLaw Journal* - Rivista di BioDiritto, 1, 2022, 17, ad esempio, ritengono tale approccio "*pienamente condivisibile*", in quanto consente una "*regolamentazione adattiva (adaptive), capace di resistere allo stress definitorio che l'evoluzione artificiale impone*". Al contrario, T. E. FROSINI, *L'orizzonte giuridico dell'intelligenza artificiale*, in *BioLaw Journal* - Rivista di BioDiritto, 1, 2022, 3-4, esprime scetticismo verso un eccesso di regolamentazione, auspicando un "*diritto minimale*" basato su principi e norme promozionali per non soffocare l'innovazione, che ritiene volano di crescita.

[35] L'argomentazione secondo cui l'approccio basato sul rischio sarebbe inadeguato a tutelare i diritti fondamentali è indirettamente supportata dalla stessa architettura dell'*AI Act* in settori ad alta sensibilità come la giustizia. P. INTURRI, S. FICHERA e A. COSTA, *La disciplina dei sistemi di intelligenza artificiale per l'amministrazione della giustizia nel Regolamento (UE) 2024/1689*, in *Lavoro Diritti Europa*, 1, 2025, 9, osservano che l'impianto normativo rivela una "*sfiducia di fondo*" verso l'IA in questo ambito, tanto da presumere per legge che tali sistemi comportino un "*rischio elevato*". Questa presunzione legale d*e facto* scavalca una valutazione del rischio caso per caso, segnalando che lo stesso legislatore europeo riconosce l'insufficienza di un mero calcolo del rischio quando sono in gioco i pilastri dello Stato di diritto.

[36] Ci riferiamo al Comitato europeo di normazione (CEN) e al Comitato europeo di normazione elettrotecnica (CENELEC), quali membri permanenti del *forum* consultivo. Il *forum* ha il compito di fornire consulenza e competenze tecniche al Consiglio per l'IA e alla Commissione Europea (art. 61 *AI Act*). E anche quali "*organizzazioni europee di normazione*" a cui la Commissione presenta le richieste per l'elaborazione di norme armonizzate. La conformità a tali norme conferisce una presunzione di conformità ai requisiti del Regolamento (art. 40 *AI Act*).

[37] La delega della definizione di *standard* tecnici a organismi privati, composti prevalentemente da attori industriali, solleva gravi problemi di legittimità democratica. Tale processo affida di fatto la concretizzazione di norme a tutela dei diritti fondamentali a soggetti portatori di interessi economici, con il rischio di una "*torsione rapace, egoistica delle regole armonizzate, ideate a vantaggio esclusivo dell'impresa*". Cfr. G. DE MINICO, *Le fonti del diritto*, cit., 87. La critica alla delega della definizione di *standard* a organismi privati trova una conferma nell'analisi del modello di co-regolamentazione dell'*AI Act*. P. INTURRI, *Intelligenza artificiale e soft law. Il ruolo dei codici di comportamento nell'Artificial Intelligence Act*, in *Nuove Autonomie*, speciale 1, 2025, 361-363, spiega come, in particolare per i "*codici di condotta*" previsti dall'art. 95, la loro elaborazione e applicazione sia rimessa "*prevalentemente alle dinamiche della prassi degli operatori del settore*". L'adesione a tali codici è volontaria e l'autore sottolinea come, in assenza di meccanismi sanzionatori forti, "*il vero incentivo all'adozione dei codici di condotta rimanga sul piano reputazionale*". Questa impostazione conferma la tesi di una tutela debole, affidata più al mercato che a vincoli giuridici cogenti. L'analisi dell'architettura dell'*AI Act* basata su una classificazione piramidale del rischio (rischio inaccettabile, elevato, limitato e minimo) è centrale

Questa debolezza sostanziale è aggravata da un paradosso interno allo stesso Regolamento europeo, che genera una sovra-regolazione formale a cui corrisponde una delega di fatto della normazione sostanziale. Da un lato, l'impianto dell'*AI Act* si fonda su principi ampi e requisiti essenziali generici, come quasi unanimente riconosciuto; dall'altro, esso delega la definizione delle specifiche tecniche cruciali a organismi a prevalente composizione privata.

L'art. 40 del Regolamento (UE) 2024/1689 affida un ruolo centrale alle "*norme armonizzate*", elaborate da enti di normazione come il Comitato europeo di normazione (CEN) e il Comitato europeo di normazione elettrotecnica (CENELEC), la cui conformità crea una vera e propria presunzione di conformità ai principali requisiti richiesti e obblighi imposti dal medesimo *AI Act*.

Analogamente, gli artt. 56 e 95 promuovono l'elaborazione di "*codici di condotta*" da parte degli stessi operatori di mercato.

Le stesse linee guida sull'ambito di applicazione degli obblighi per i modelli di IA per scopi generali della Commissione[38] chiariscono che l'adesione a tali codici, se valutati come adeguati, rappresenta per i fornitori un modo diretto per dimostrare la conformità, beneficiando di una maggiore fiducia da parte della Commissione e potenzialmente vedendo ridotto il numero di richieste di informazioni. Si perfeziona così una delega che sposta il baricentro normativo dal legislatore democratico a soggetti privati, i cui *standard* sono guidati da logiche di mercato e non primariamente dalla tutela dei diritti.

Su questo impianto già critico si innesta il disegno di legge italiano, che amplifica il fenomeno della sovra-regolazione formale attraverso il cosiddetto *gold-plating*. Disposizioni come l'art. 1, comma 1, e l'art. 3, commi 1, 2 e 3, del disegno di legge italiano si limitano a ribadire a livello nazionale principi - quali la trasparenza e la dimensione antropocentrica - già giuridicamente vincolanti in virtù del primato del Regolamento. Questa duplicazione, riconosciuta implicitamente dallo stesso art. 1, comma 2, del disegno di legge, non aggiunge forza cogente ma crea un apparato legislativo sovrabbondante che, nel replicare la genericità del modello europeo, ne eredita e ne amplifica le debolezze, affidando la tutela dei diritti a un sistema stratificato ma privo (ancora) di un'effettiva prescrittività.

L'approccio basato sul rischio si dimostra così già *prima facie* inefficace e illusoriamente flessibile. La classificazione del rischio si rivela rigida e statica, incapace di adattarsi alla natura dinamica e contestuale dei pericoli stessi posti dall'IA.

L'Allegato III dell'*AI Act*, che elenca i settori ad "*alto rischio*" di pregiudicare la salute e la sicurezza o i diritti fondamentali delle persone - tra cui l'amministrazione della giustizia, l'accesso a servizi pubblici e privati essenziali, l'istruzione e il lavoro - si basa su una concezione del "*danno*" prevalentemente individuale, immediato e materiale.

---

anche nel contributo di R. RAZZANTE, *AI e tutela dei diritti fondamentali*, in *dirittifondamentali.it*, 1, 2024, 133-157, che ne illustra la struttura. La sua riflessione sulla compatibilità di tale impianto con la Costituzione italiana offre un utile contrappunto. Laddove il nostro saggio critica tale approccio come una "*de-costituzionalizzazione*" dei diritti, l'autore esplora come i principi costituzionali (artt. 2, 3, 21, 32) debbano fungere da guida per l'applicazione della nuova normativa. Di particolare interesse è la sua adesione a una prospettiva "*algoretica*", che postula la necessità di interiorizzare l'elemento etico nell'algoritmo, un concetto che converge con la nostra proposta di un "*costituzionalismo algoritmico sostanziale*" che renda la compatibilità costituzionale un requisito di progettazione.

[38] *Guidelines on the scope of the obligations for general-purpose AI models established by Regulation (EU) 2024/1689*, pubblicate il 18 luglio 2025, insieme al *Code of practice for General-Purpose AI Models*, in *https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai*.

Questo modello è strutturalmente cieco ai rischi che sono, al contempo, i più insidiosi per le democrazie costituzionali: quelli sistemici, a lungo termine e immateriali, come si analizzerà in seguito[39].

Gli stessi orientamenti della Commissione Europea sul divieto di punteggio sociale (art. 5, par. 1, lett. c)[40], ad esempio, ne confermano l'impostazione individualistica e correttiva. Il divieto, infatti, non è assoluto, ma scatta solo in presenza di due scenari specifici: quando il punteggio comporta un trattamento pregiudizievole in contesti sociali "*non collegati*" a quelli della raccolta dati, oppure quando tale trattamento è "*ingiustificato o sproporzionato*" rispetto al comportamento sociale.

L'analisi della Commissione si concentra interamente sulla *fairness* del processo a livello individuale - ad esempio, l'uso di dati pertinenti per la valutazione del merito creditizio è considerato lecito - ignorando completamente il rischio sistemico che una società basata sul punteggio generalizzato comporta. La questione non è solo se il singolo punteggio sia "*giusto*", ma l'impatto collettivo di una sorveglianza pervasiva che incentiva il conformismo e mina alla radice la fiducia e la coesione sociale, erodendo proprio quei processi democratici che il modello basato sul rischio si dimostra non del tutto capace di proteggere.

Anche per i sistemi classificati ad "*alto rischio*", i meccanismi di mitigazione e di conformità si rivelano deboli. La scelta di affidare la valutazione di conformità, nella maggior parte dei casi, a un meccanismo di *self-compliance* da parte del fornitore, integrato da *standard* tecnici elaborati da organismi privati, costituisce una *delega di fatto* della tutela dei diritti[41] a soggetti portatori di interessi economici, con potenziali, ma pur sempre evidenti e insanabili conflitti[42].

A ciò si aggiunga la previsione di strumenti come la "*valutazione d'impatto sui diritti fondamentali*" (*Fundamental Rights Impact Assessment* o FRIA) - prevista dall'art. 27 solo per

[39] Cfr. F. DONATI, *La protezione dei diritti fondamentali*, cit., 17, per il quale la scelta di un approccio basato sul rischio, pur essendo "*sostanzialmente condivisibile*" per bilanciare tutela e innovazione, presenta notevoli criticità. In particolare, la definizione *ex ante* delle categorie di rischio da parte del legislatore rischia di rivelarsi "*inadeguata alla prova dei fatti o comunque obsoleta per la rapidità dell'evoluzione tecnologica*", delegando di fatto il bilanciamento alla Commissione Europea.

[40] Orientamenti della Commissione relativi alle pratiche di intelligenza artificiale vietate ai sensi del regolamento (UE) 2024/1689 (regolamento sull'IA), pubblicati il 4 febbraio 2025 in *https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act*.

[41] G. DE MINICO, *Le fonti del diritto*, cit., 88, dove si argomenta che rimettere a uno *standard* tecnico privato la definizione del rischio relativo a equità e diritti fondamentali "*non è un'operazione neutrale perché comporta una preferenza per una logica specifica e un insieme di priorità in luogo di altre pure possibili*", delegando di fatto scelte politiche a organismi privi della necessaria legittimazione. L'affidamento su meccanismi di autovalutazione da parte dei fornitori, come la *self-compliance* o la *Fundamental Rights Impact Assessment*, è uno dei maggiori punti deboli dell'*AI Act*. La coincidenza soggettiva tra controllato e controllante "*vanifica la funzione cautelativa della valutazione*", creando una mera "*parvenza di liceità*" che non offre garanzie di neutralità e obiettività. F. DONATI, *La protezione dei diritti fondamentali*, cit., 19, secondo cui una delle principali criticità del Regolamento risiede proprio nel sistema di controllo, che affida la verifica della conformità principalmente ai fornitori, limitando l'intervento di organismi terzi notificati. L'efficacia del sistema di tutela è quindi "*rimesso in buona parte alle scelte degli Stati membri*" sulla creazione di autorità di controllo "*efficienti e ben preparate*". Cfr. E. CIRONE, *L'AI Act e l'obiettivo (mancato?)*, cit., 14, la quale, pur riconoscendo l'introduzione della FRIA come un passo avanti, rileva come essa "*non sembra essere stata pienamente integrata nel quadro normativo del regolamento, mettendo ancor più in luce come questa integrazione sia stata frutto di un compromesso politico, nell'ambito del quale molti degli elementi chiave per una strutturata e adeguata valutazione di impatto sui diritti fondamentali sono stati omessi*".

[42] Per una critica all'affidamento alla standardizzazione tecnica per la tutela dei diritti fondamentali, v. F. DONATI, *La protezione dei diritti fondamentali*, cit., 10 ss.; E. CIRONE, *L'*AI Act *e l'obiettivo (mancato?)*, cit.

una ristretta cerchia di utilizzatori (principalmente enti pubblici e operatori bancari/assicurativi) - che, essendo redatta dallo stesso fornitore, rischia di vanificare la sua funzione cautelativa, creando una mera "*parvenza di liceità*"[43], e configurandosi più come un adempimento burocratico che come un effettivo strumento di controllo indipendente e sostanziale[44].

La giustificazione secondo cui un approccio più stringente soffocherebbe l'innovazione si fonda su una visione parziale e non bilanciata del progresso. Non tiene conto dei costi nascosti e delle esternalità negative che un'innovazione non governata impone alla collettività. Un'innovazione che procede al costo di erodere le fondamenta della fiducia democratica o di accelerare la crisi ambientale non è un progresso netto, ma un trasferimento di costi dal settore tecnologico alla società nel suo complesso.

Un quadro regolatorio più esigente, basato su principi di un costituzionalismo algoritmico sostanziale, non soffocherebbe l'innovazione, ma la orienterebbe verso direzioni socialmente più desiderabili e sostenibili.

Questa prospettiva trova un solido ancoraggio costituzionale se si considera lo sviluppo e l'impiego di sistemi di IA come una moderna e pervasiva forma di iniziativa economica privata, ricadente nell'ambito dell'art. 41 Cost.

Come ben evidenziato in dottrina, la libertà d'impresa non è mai assoluta, ma è subordinata al rispetto di limiti invalicabili che, nella loro essenza, proteggono la persona. In particolare, si sostiene che il trinomio "*sicurezza, libertà, dignità umana*" vada letto come un trinomio inscindibile, un "*climax ascendente*" che evoca i valori personalistici dell'art. 2 Cost. e li eleva "*a parametro e fine di un uso responsabile*" della libertà economica[45].

Applicando questo schema all'era digitale, il principio del costituzionalismo algoritmico sostanziale non è altro che l'attuazione del disegno dei Costituenti: così come l'attività d'impresa deve essere orientata alla dignità umana, allo stesso modo l'innovazione tecnologica deve essere progettata per essere intrinsecamente compatibile con i diritti fondamentali, trasformando il limite costituzionale da vincolo esterno a principio di progettazione.

Un quadro di tal fatta stimolerebbe inoltre la ricerca e lo sviluppo di un'IA affidabile, trasparente ed equa, creando un vantaggio competitivo per l'Europa basato non solo sulla *performance* economica, ma anche sulla compatibilità con i valori democratici.

---

[43] G. DE MINICO, *Giustizia e intelligenza artificiale*, cit., 87. L'autrice critica aspramente il meccanismo della *self-compliance* e della valutazione d'impatto autodichiarata, evidenziando come la coincidenza soggettiva tra controllato e controllante comprometta l'obiettività del sindacato, che risulta "*strutturalmente inadatto a offrire quelle garanzie di neutralità necessarie a rassicurare i terzi*".

[44] O. POLLICINO, *Regolazione e innovazione tecnologica*, cit., 165 ss., critica aspramente l'affidamento sull'autovalutazione da parte dei *deployer*, come nel caso della *Fundamental Rights Impact Assessment* (FRIA), che, senza un'adeguata supervisione indipendente, rischia di trasformarsi in un mero adempimento burocratico, minando l'efficacia delle tutele previste. Anche la dottrina più recente solleva perplessità sul meccanismo della valutazione d'impatto sui diritti fondamentali (FRIA), notando come l'obbligo sia imposto solo a una ristretta cerchia di soggetti. Come osserva, infatti, F. DONATI, *La protezione dei diritti fondamentali*, cit., 18, "*se la valutazione di conformità e i controlli previsti per i sistemi di IA ad alto rischio non sono ritenuti in grado di assicurare in linea generale un elevato livello di protezione dei diritti fondamentali, allora la valutazione d'impatto andrebbe estesa a tutti gli utilizzatori di tali sistemi*".

[45] G. A. FERRO, *Osservazioni sul principio di "sicurezza" nella Costituzione italiana e sulle sue valenze quale limite alla libertà di iniziativa economica*, in F. La Rosa (a cura di), *Costituzione, legalità e aziende. Raccolta interdisciplinare di saggi del progetto formativo A-CISCO*, FrancoAngeli, Milano, 2023, 709.

In definitiva, la critica all'approccio basato sul rischio non nega la necessità di proporzionalità o il valore dell'innovazione. Al contrario, essa sostiene che, nel caso specifico dell'IA, questo modello rappresenta una disapplicazione di tali principi.

In definitiva, l'approccio europeo, e in prospettiva italiano, non sembra genuinamente proporzionato perché tratta i diritti fondamentali come rischi negoziabili. Non è efficace perché è cieco ai pericoli più gravi e sistemici. E la sua difesa in nome dell'innovazione ignora i costi enormi che questa stessa innovazione, se non governata, scarica sulla società. I rischi per l'impianto costituzionale superano di gran lunga i benefici attesi da un modello regolatorio così debole e meramente correttivo.

4. Le dimensioni sistemiche e collettive trascurate.

L'approccio dell'*AI Act*, focalizzato sul rischio del singolo sistema e sul danno individuale, si dimostra strutturalmente incapace di cogliere e affrontare i rischi di natura sistemica.

Si tenta di risolvere un problema di proprietà emergenti di un intero *ecosistema* tecnologico con strumenti pensati per rischi contenuti e individualizzabili[46].

È pur vero che il legislatore europeo ha dedicato un'attenzione specifica ai modelli di IA per finalità generali (GPAI) che presentano "*rischi sistemici*", identificati principalmente attraverso una soglia di calcolo per l'addestramento superiore a 10^{25} FLOP. Per questi modelli, l'art. 55 dell'*AI Act* impone obblighi aggiuntivi, dettagliati nel recentissimo *Code of Practice for General-Purpose AI Models - Safety and Security Chapter*. Tale codice di condotta impegna i firmatari ad adottare un "*Safety and Security Framework*" all'avanguardia per la gestione continua dei rischi sistemici lungo l'intero ciclo di vita del modello. Il codice stesso elenca esplicitamente tra i rischi da considerare quelli per i "*processi democratici*", la "*salute pubblica*", la "*sicurezza*" e i "*diritti fondamentali*".

Tuttavia, anche questo approccio rafforzato rimane ancorato alla logica della gestione del rischio. Esso si basa su processi di autovalutazione (*risk assessment*), mitigazione e segnalazione di "*incidenti gravi*" (*serious incident reporting*), che sono per loro natura correttivi e procedurali. La tutela, ancora una volta, è affidata alla capacità del fornitore di identificare, analizzare e mitigare i rischi, senza però imporre vincoli strutturali alla progettazione della tecnologia per prevenire tali rischi alla fonte.

Tra le molteplici dimensioni sistemiche trascurate, due comunque emergono con particolare urgenza per la loro portata costituzionale: l'impatto ambientale e l'impatto sui processi democratici. Questi temi, che trascendono la dimensione della tutela individuale del diritto, rivelano l'inadeguatezza del modello basato sul rischio non solo a livello individuale, ma anche collettivo e intergenerazionale.

In primo luogo, il prevedibile impatto ambientale: l'enorme consumo di risorse energetiche e idriche richiesto dai *data center* per l'addestramento e il funzionamento dei modelli di IA più avanzati rappresenta una seria minaccia per la sostenibilità e la giustizia intergenerazionale, un tema quasi completamente trascurato dall'*AI Act* e dal disegno di

[46] La complessità dell'ecosistema dell'IA, e la conseguente difficoltà di regolarlo, è ben illustrata dalla distinzione tra "*modello di IA*" e "*sistema di IA*". Come spiegano P. INTURRI, S. FICHERA e A. COSTA, *La disciplina dei sistemi di intelligenza artificiale*, cit., 7-8, un modello (es. un GPAI) è un algoritmo addestrato, mentre un sistema è l'insieme di componenti preordinato a un compito specifico. La disciplina dell'*AI Act* si differenzia: gli obblighi per i modelli GPAI sono "*orizzontali*" e indipendenti dall'uso, mentre quelli per i sistemi dipendono dal contesto di utilizzo e dal livello di rischio. Questa architettura normativa, che frammenta la responsabilità lungo la catena del valore, rafforza la tesi secondo cui un approccio focalizzato solo sul "*danno*" finale del sistema è insufficiente, poiché ignora le problematiche sistemiche che nascono già a livello del modello sottostante.

legge italiano[47]. Questo impatto non è un'esternalità secondaria, ma una questione di rilevanza costituzionale, che intercetta direttamente l'art. 9 Cost., come modificato dalla legge costituzionale 11 febbraio 2022, n. 1, il quale impegna la Repubblica alla tutela dell'ambiente, della biodiversità e degli ecosistemi, "*anche nell'interesse delle future generazioni*" [48].

L'*AI Act*, si concentra sul rischio del singolo sistema. Un modello di IA, valutato individualmente, non presenta un rischio ambientale diretto. Il rischio emerge dalla somma dei consumi di migliaia di *data center* che compongono l'infrastruttura globale dell'IA.

Un approccio regolatorio sistemico, al contrario, affronterebbe il problema alla radice, imponendo ad esempio *standard* di efficienza energetica e idrica per l'intero settore, promuovendo la localizzazione dei *data center* in aree con minori criticità ambientali e imponendo obblighi di trasparenza radicale sui consumi.

Il secondo rischio sistemico è il potenziale impatto sui processi democratici: l'uso pervasivo del *micro-targeting* politico, che consente di inviare messaggi personalizzati e finemente calibrati a specifici segmenti dell'elettorato, basandosi su dati degli utenti raccolti *online*, potenziato dall'avvento dell'IA generativa, capace di creare *deepfake* altamente realistici, di generare automaticamente enormi volumi di contenuti di propaganda personalizzati e di impiegare *chatbot* per simulare conversazioni politiche su larga scala, potrebbe consentire forme sofisticate di manipolazione del consenso elettorale, un rischio che trascende la classificazione del singolo sistema e investe il principio supremo democratico[49].

5. Verso un approccio sostanziale e preventivo: la primazia della persona e il diritto all'interazione umana.

Diverse alternative al modello dell'*AI Act* sono ipotizzabili. Sebbene ciascuna presenti elementi di interesse, tutte condividono un limite fondamentale: rimangono, in larga misura, approcci correttivi e procedurali, che non sembrano affrontare adeguatamente il problema strutturale dell'opacità algoritmica alla sua radice.

Un primo modello alternativo si potrebbe basare sul rafforzamento dei regimi di responsabilità civile. Si potrebbe ipotizzare un regime di responsabilità civile aggravata, o addirittura oggettiva, per i produttori di sistemi di IA ad alto rischio. Questo approccio, che trova un parallelo nella responsabilità per attività pericolose (art. 2050 c.c.) o per danno da

[47] La preoccupazione per l'impatto sistemico dell'IA sui processi democratici è ampiamente condivisa. T. GROPPI, *Alle frontiere dello Stato costituzionale*, cit., 679, identifica una minaccia diretta alla sovranità popolare, derivante dalla capacità degli algoritmi di influenzare il consenso tramite la profilazione degli utenti e la creazione di *bubble democracy*, che favoriscono "*incomunicabilità, polarizzazione, quando addirittura non si giunga fino all'incitamento all'odio*". Allo stesso modo, C. CASONATO, *Costituzione e intelligenza artificiale*, cit., 4-5, avverte che un uso incontrollato dell'IA nella propaganda politica rischia di svuotare "*dall'interno della sovranità popolare*", erodendo il *marketplace of ideas* su cui si fonda una libera competizione pubblica. M. TOMASI, *Intelligenza artificiale, sostenibilità e responsabilità intergenerazionali: nuove sfide per il costituzionalismo?*, in *Rivista AIC*, 4, 2024, 49 ss.

[48] L'insufficiente attenzione dell'*AI Act* agli aspetti di sostenibilità ambientale è stata evidenziata come una grave lacuna. Affidare un tema di tale portata, con impatti enormi in termini di consumo energetico e idrico, a meri "*codici di condotta volontari*" appare in forte contraddizione con la politica del *Green Deal* europeo. Sul punto, F. DONATI, *La protezione dei diritti fondamentali*, cit., 19.

[49] Cfr. A. SIMCHON, M. EDWARDS e S. LEWANDOWSKY, *The persuasive effects of political microtargeting in the age of generative artificial intelligence*, in *PNAS Nexus*, 29 gennaio 2024. Sul punto, si veda G. MAIRA, *Intelligenza umana e intelligenza artificiale*, in *federalismi.it*, 7, 2021, XIV, che ricorda come, nel caso *Cambridge Analytica*, la raccolta di informazioni personali di milioni di utenti *Facebook* "*è stata in grado di tracciare dei profili psicologici con cui produrre il rafforzamento delle opinioni e influenzare il voto politico*".

prodotti difettosi, avrebbe il pregio di facilitare il risarcimento per le vittime, invertendo l'onere della prova e incentivando i produttori a investire in sicurezza.

Il disegno di legge italiano, all'art. 24, comma 5, lett. d), sembra muoversi in questa direzione delegando, tra l'altro, il Governo a prevedere "*nei casi di responsabilità civile,… strumenti di tutela del danneggiato, anche attraverso una specifica regolamentazione dei criteri di ripartizione dell'onere della prova, tenuto conto della classificazione dei sistemi di intelligenza artificiale e dei relativi obblighi come individuati dal regolamento (UE) 2024/1689*". Tuttavia, la natura del modello civilistico è intrinsecamente *ex post*: interviene per "*compensare*" il danno dopo che si è verificato, non per prevenirlo. Inoltre, si scontra con le enormi difficoltà probatorie legate alla *black box* e alla frammentazione della catena di responsabilità tra sviluppatori, fornitori di dati, e utilizzatori.

Un secondo paradigma, comunque complementare a quello della responsabilità civile, si potrebbe concentrare sul potenziamento dei poteri di *audit* e di ispezione delle autorità di controllo. Si potrebbero conferire alle Autorità Nazionali Competenti (NCA)[50], come quelle previste dall'*AI Act*, poteri di indagine molto più penetranti, inclusi l'accesso al codice sorgente, ai dati di addestramento e la capacità di condurre *audit* tecnici approfonditi. Questo aumenterebbe senza dubbio la capacità di *enforcement* e di controllo.

Ciononostante, anche questo modello incontra il limite invalicabile dell'opacità. Verificare un sistema di *deep learning* complesso e in continua evoluzione è un'impresa tecnicamente ardua, se non impossibile. Il rischio è che l'*audit* si trasformi in un controllo formale sulla documentazione e sulle procedure, senza riuscire a penetrare la logica sostanziale del "*ragionamento*" della macchina.

Un terzo approccio, anch'esso complementare ai precedenti, potrebbe essere quello della generalizzazione dei sistemi di certificazione obbligatoria da parte di terzi indipendenti.

In questo modello, i sistemi di IA ad alto rischio non potrebbero essere immessi sul mercato senza aver prima ottenuto una certificazione da un organismo accreditato, che verifichi non solo la conformità tecnica ma anche quella ai diritti fondamentali. Questo meccanismo, parzialmente previsto dall'*AI Act* per alcuni sistemi, potrebbe essere generalizzato e reso più stringente. Il vantaggio sarebbe l'introduzione di un vaglio indipendente *ex ante*.

La debolezza risiede nel rischio che la certificazione diventi un adempimento burocratico, un "*bollino*" di conformità che non garantisce un'effettiva aderenza sostanziale ai principi costituzionali, specialmente di fronte a sistemi che si adattano e cambiano dopo la loro immissione sul mercato.

Si apprezza in particolare il contributo di chi ha originalmente valorizzato il momento cruciale dell'acquisto dei sistemi di IA da parte della pubblica amministrazione[51]. In questa prospettiva, il contratto pubblico non è un mero atto di compravendita, ma si configura come un potente strumento di regolazione preventiva.

Tuttavia, come evidenziato dallo stesso autore, questo potenziale è oggi neutralizzato dall'insufficienza dell'architettura normativa dell'*AI Act*, le cui regole si mantengono a un livello di principi troppo generici per essere efficaci in sede di gara.

[50] Sulle quali si rimanda al seguito del testo, in particolare al par. 5.

[51] Cfr. G. F. LICATA, *Intelligenza artificiale e contratti pubblici*, cit. L'analisi di Licata evidenzia in modo puntuale come la contrattualistica pubblica, pur essendo uno strumento potenzialmente potente per orientare lo sviluppo di un'IA "*responsabile*", rischi di essere depotenziata dalla genericità dei principi dell'*AI Act* e dalla supremazia tecnica degli operatori privati, confermando la necessità di una regolazione esterna più stringente e sostanziale.

L'amministrazione che intende acquistare un'IA "*costituzionalmente orientata*" non trova nel Regolamento europeo gli *standard* tecnici e giuridici sufficientemente dettagliati per tradurre tale esigenza in precise clausole contrattuali, rischiando così di subire le condizioni e gli *standard* qualitativi imposti dagli operatori di mercato. Di qui, la tesi sulla necessità di un deciso implemento della regolazione amministrativa esterna, che vada a definire *standard* concreti e vincolanti.

Si delinea così l'esigenza di una tutela anticipata ed effettiva da attuarsi già al momento dell'acquisto, al fine di strutturare sin da quella fase un'intelligenza artificiale qualitativamente adeguata e conforme all'interesse pubblico.

Questo approccio, focalizzato sul contratto pubblico di acquisto dei sistemi di IA, converge pienamente con la critica più generale qui mossa all'approccio basato sul rischio: esso conferma che una tutela meramente *ex post* è inadeguata e che il governo della tecnologia richiede meccanismi di intervento preventivi e sostanziali, capaci di plasmare l'innovazione prima che essa produca i suoi effetti sulla società.

Muovendoci, dunque, nella direzione dell'anticipazione della tutela dei diritti fondamentali, un'alternativa credibile all'approccio basato sul rischio non può che fondarsi sui principi costituzionali che affermano la centralità e l'inviolabilità della persona umana.

L'art. 2 della Costituzione italiana, nel riconoscere e garantire i diritti inviolabili dell'uomo, sia come singolo sia nelle formazioni sociali ove si svolge la sua personalità, e l'art. 3, comma 2, che impegna la Repubblica a rimuovere gli ostacoli di ordine economico e sociale che impediscono il pieno sviluppo della persona umana, forniscono il solido fondamento per una "*riserva di umanità*" da rafforzare e considerare come invalicabile dalla tecnologia[52].

La centralità della persona, che fonda questa "*riserva di umanità*", non è un principio astratto, ma un criterio che orienta l'intero ordinamento, inclusa la sua "*Costituzione economica*".

Lo stesso trinomio "*sicurezza, libertà, dignità umana*", posto dall'art. 41 Cost. come limite all'iniziativa economica, non serve a tutelare un generico interesse collettivo, ma, come bene argomentato in dottrina[53], a garantire proprio la dimensione personalistica. Nella sua lettura, la Costituzione è ragionevole ritenere connetta a tutti e tre i termini l'attributo "*umana*", proprio per "*evocare integralmente i valori personalisti racchiusi nell'art. 2 Cost.*". Se anche il potere economico è costituzionalmente finalizzato alla piena realizzazione della persona, a maggior ragione deve esserlo il potere tecnologico, quale frutto del primo.

La proposta di un "*diritto all'interazione umana*" si rivela, dunque, non un'interpretazione audace, ma la coerente applicazione di un principio cardine del nostro ordinamento: ogni forma di potere deve arrestarsi di fronte alla dignità della persona, che ne rappresenta il fine ultimo e il limite invalicabile. Su queste basi, e accogliendo le tesi innovative elaborate da autorevole dottrina, è possibile, innanzi tutto, confermare i lineamenti di un vero e proprio "*diritto all'interazione umana*" come diritto fondamentale autonomo[54].

La Costituzione italiana delinea un ordinamento il cui baricentro non è (o comunque non è soltanto) lo Stato-apparato, ma prima di tutto la persona umana. Questa impostazione fornisce la base per argomentare la necessità di un rapporto non meramente automatizzato tra cittadino e poteri pubblici. Il principio personalista di cui all'art. 2 Cost. è il fondamento di un rapporto con i pubblici poteri non interamente mediato da sistemi automatizzati, che tratti l'individuo non soltanto come un insieme di dati da processare, ma che riconosca e

[52] C. SAGONE, *Efficientamento della giustizia e intelligenza artificiale*, in *Rivista AIC*, 1, 2024, 309.
[53] G. A. FERRO, *Osservazioni sul principio di "sicurezza"*, cit., 712.
[54] G. SCACCIA e A. MONORITI, *Quali spazi per un diritto all'interazione umana?*, cit., 1 ss.

garantisca la capacità del cittadino di partecipare attivamente alle "*formazioni sociali*", prima fra tutte la comunità politica.

Lo stesso adempimento dei doveri inderogabili di solidarietà presuppone un modello relazionale e di dialogo umano. La rimozione degli ostacoli di cui all'art. 3, comma 2, Cost. richiede un'azione amministrativa basata su un'istruttoria approfondita di ogni rapporto, che passa necessariamente attraverso l'ascolto e il dialogo tra persone fisiche.

Il principio di sovranità popolare di cui all'art. 1 Cost., la garanzia dell'accesso ai pubblici uffici in condizioni di eguaglianza di cui all'art. 51 Cost., e la promozione del principio di sussidiarietà orizzontale che favorisce "*l'autonoma iniziativa dei cittadini, singoli e associati*" di cui all'art. 118, ultimo comma, Cost., disegnano un rapporto collaborativo e dialogico tra amministrazione e cittadini.

Ancora, un'amministrazione che opera secondo "*buon andamento*" è quella che costruisce un rapporto di fiducia con i cittadini, garantendo la legittimità e la trasparenza del proprio operato. Tale fiducia non è un concetto meramente tecnico, ma una categoria giuridica e semantica che implica aspettativa, continuità e responsabilità. Da essa discende la fidelizzazione, ossia la capacità delle istituzioni di mantenere nel tempo un legame stabile con la collettività. La fiducia rimanda al patto, all'affidamento reciproco e al riconoscimento interpersonale. Essa implica non solo affidabilità tecnica, ma anche un rapporto umano fatto di empatia e responsabilità morale. È proprio questa dimensione relazionale che consente alla fiducia di costituire il fondamento del legame tra istituzioni e cittadini. L'interazione umana, l'ascolto e la comunicazione ne diventano componenti essenziali.

L'imparzialità stessa è tutelata quando una decisione è riconducibile a un funzionario specifico, che può renderne conto e assumersene la responsabilità. La Legge 7 agosto 1990, n. 241, attuando i principi di partecipazione, di trasparenza e il diritto costituzionale di difesa (art. 24 Cost.), realizza uno schema costituzionalmente orientato di procedimento amministrativo decisionale fondato sul dialogo umano, tramite la personificazione della responsabilità nel responsabile del procedimento, persona fisica, e il modello della partecipazione che va dalla comunicazione di avvio del procedimento, passa dal diritto di intervento e dal preavviso di rigetto, quali strumenti funzionali al dialogo umano stesso.

Gli stessi art. 24 e 111 Cost. sanciscono la facoltà di partecipare alla formazione di una decisione e il contraddittorio quale "*metodo*" di confronto dialettico, non meramente limitati al "*giusto processo*", ma tendenzialmente espansivi verso tutti i rapporti tra cittadini e pubblici poteri.

Sebbene non esplicitamente codificato, emerge dunque dall'ordinamento costituzionale (e primario) italiano un chiaro e inequivoco diritto all'interazione umana con i pubblici poteri.

Questo diritto ha natura strumentale: è il mezzo indispensabile per rendere effettivi altri diritti e principi fondamentali in un contesto di crescente automazione. È lo strumento per garantire la dignità della persona (art. 2 Cost.), per perseguire l'uguaglianza sostanziale (art. 3 Cost.), per assicurare il diritto di difesa (art. 24 Cost.), per dare sostanza al "*giusto procedimento*" (L. n. 241/1990), per attuare il "*giusto processo*" attraverso il contraddittorio tra le parti e per realizzare un "*buon andamento*" che sia anche umano e relazionale (art. 97 Cost.).

Tale diritto va nettamente distinto dalla mera "*sorveglianza umana*" (*human-in-the-loop*), prevista dall'art. 14 dell'*AI Act* e dal principio di "*non esclusività*" della decisione algoritmica[55], enucleato dalla giurisprudenza amministrativa[56].

La sorveglianza umana, così come delineata dal Regolamento, si configura come un controllo a valle, un presidio funzionale che interviene sull'*output* del sistema. Il recentissimo *Code of Practice - Safety and Security Chapter*, ad esempio, la considera un fattore contestuale nella valutazione dei rischi sistemici. Il *Code of Practice - Transparency Chapter*", a sua volta, si concentra sull'obbligo di fornire documentazione e informazioni ai fornitori a valle ("*downstream providers*"), rafforzando l'idea di una catena di responsabilità procedurale, ma non affrontando la qualità della decisione finale.

Gli orientamenti della Commissione[57], nel dettagliare le deroghe all'uso dell'identificazione biometrica remota da parte delle forze dell'ordine, chiariscono la natura di questo intervento umano: nessuna decisione con effetti giuridici negativi può essere presa "*unicamente sulla base dell'*output *del sistema*", e l'identificazione deve essere "*verificata e confermata separatamente da almeno due persone fisiche*".

Il disegno di legge italiano, all'art. 3, comma 3, ha il merito di enunciare la sorveglianza umana come un principio cardine e irrinunciabile per lo sviluppo e l'applicazione dell'IA, oltre i sistemi ad alto rischio, e pur tuttavia, per la sua concreta applicazione, esso si affida completamente e implictamente al Regolamento UE.

La sorveglianza umana, dunque, si configura come un controllo funzionale, meramente formale e successivo (*ex post*), sull'*output* generato dalla macchina[58].

Il diritto all'interazione umana, al contrario, è un diritto sostanziale a un rapporto dialogico, empatico e comprensivo con un altro essere umano *durante* il procedimento decisionale, anche algoritmico che incide sulla sfera giuridica del destinatario[59].

---

[55] Sulla codificazione del principio di "*non esclusività della decisione algoritmica*" e del c.d. "Human in the Loop", si veda D. U. GALETTA, *Digitalizzazione, Intelligenza artificiale e Pubbliche Amministrazioni: il nuovo Codice dei contratti pubblici e le sfide che ci attendono*, in *federalismi.it*, 12, 2023, XI. L'autrice, tuttavia, ne sottolinea il carattere "*del tutto ridondante*" e ne evidenzia i limiti pratici, legati alla difficoltà psicologica per il funzionario di discostarsi dal risultato proposto dalla "*onnipotente "macchina'*"". Ancora più critica, sempre D. U. GALETTA, *Human-stupidity-in-the-loop?*, cit., XIII, la quale teme che l'interazione umana, a causa dei *bias* cognitivi e della tendenza a validare le proprie convinzioni, possa trasformare l'Intelligenza Artificiale in un "*effetto moltiplicatore delle imperfezioni*" umane, creando uno scenario di "Human-stupidity-in-the-loop".

[56] Consiglio di Stato, Sez. VI, 13 dicembre 2019, n. 8472, che ha enucleato il principio di "*non esclusività della decisione algoritmica*".

[57] Orientamenti della Commissione relativi alle pratiche di intelligenza artificiale vietate ai sensi del regolamento (UE) 2024/1689 (regolamento sull'IA), pubblicati il 4 febbraio 2025 in *https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act*.

[58] La necessità di un intervento umano non è un mero gesto simbolico (*rubber-stamping*), ma deve consistere in un "*controllo significativo*" svolto da una persona dotata di "*sufficiente autorità e adeguata competenza per modificare la decisione meccanica*". Sul punto G. DE MINICO, *Giustizia e intelligenza artificiale*, cit., 97.

[59] Il principio di "*non esclusività*", elaborato dalla giurisprudenza amministrativa, impone che la decisione algoritmica sia sempre soggetta a un intervento umano qualificato. Esso richiede "*un contributo umano capace di controllare, validare ovvero smentire la decisione automatica*", secondo il modello *HITL* (*human in the loop*). Cfr. C. NAPOLI, *Algoritmi, intelligenza artificiale e formazione della volontà pubblica*, cit., 338.La distinzione tra l'impiego dell'IA come strumento di ausilio e la sua inammissibile sostituzione al giudicante è ben delineata in F. DONATI, *Intelligenza Artificiale e giustizia*, cit., 430 ss., dove si ipotizzano usi virtuosi della tecnologia per la ricerca di precedenti, la gestione di cause seriali o il supporto in valutazioni tecniche. La proposta di un "*diritto fondamentale all'interazione umana*" trova un importante parallelo nella dottrina che chiede di non sostituire, ma di assistere l'uomo con l'IA. C. CASONATO, *Costituzione e intelligenza artificiale*, cit., 10-12, ad esempio, propone un "*diritto ad essere destinatari di decisioni che siano il risultato di un processo in cui sia*

L'interazione umana non è un semplice meccanismo di controllo, ma il veicolo di valori non computabili e non riducibili a dati: l'empatia, il riconoscimento della dignità altrui, la comprensione del contesto, la capacità di ponderare interessi non formalizzabili[60]. Questi elementi sono essenziali per decisioni giuste e umane, specialmente in settori ad alta densità relazionale come la sanità[61], il lavoro[62], la giustizia[63] e l'accesso ai servizi sociali, come lo stesso legislatore italiano sembra intenzionato a volere riconoscere in linea di principio[64].

La "*riserva di umanità*" andrebbe così intesa, ancorata e potenziata, per potere operare come un limite costituzionale implicito alla delega di funzioni pubbliche a sistemi automatizzati ad alto rischio, in particolare quando tali funzioni implichino l'esercizio di discrezionalità.

L'uso dell'IA nella pubblica amministrazione e nell'attività giudiziaria, come previsto dagli artt. 14 e 15 del d.d.l. A.S. n. 1146-B, deve essere e rimanere rigorosamente strumentale e di supporto, senza mai poter sostituire la valutazione ponderata (e la responsabilità finale) dell'essere umano[65].

Pur apprezzandosi l'impianto generale antropocentrico del disegno di legge italiano, è necessario evidenziarne le ambiguità: l'enfasi sui "*principi*" rischia di rimanere una mera dichiarazione programmatica se non attraverso il (collegamento al) diritto fondamentale all'interazione umana come pretesa soggettiva, azionabile e giustiziabile.

L'automazione integrale dei rapporti tra cittadino e potere pubblico, come nell'erogazione di prestazioni sociali o nell'accesso a diritti, crea una nuova e insidiosa forma di disuguaglianza. I cittadini con maggiori competenze digitali e risorse economiche per contestare le decisioni algoritmiche si trovano in una posizione di vantaggio. Al contrario, i soggetti più vulnerabili - anziani, persone con disabilità, migranti, persone a basso reddito -

---

presente una significativa componente umana", fondato sulla necessità di individuare la titolarità e la responsabilità della decisione, nonché di preservare qualità umane non replicabili come l'empatia e la comprensione del contesto. Questo va oltre il mero controllo umano sull'output finale, come confermato dalla giurisprudenza amministrativa che ha elaborato il principio di "*non esclusività*" della decisione algoritmica, richiedendo un "*intervento umano qualificato (di controllo o di verifica) nel processo decisionale*". Il saggio di C. EQUIZI, *Intelligenza artificiale: profili di opportunità e di criticità nella irrinunciabile tutela dei diritti fondamentali*, in *dirittifondamentali.it*, 1, 2024, 313-331, rafforza la necessità di un presidio umano invalicabile nelle decisioni automatizzate. Analizzando casi concreti di discriminazione algoritmica, come la vicenda statunitense "*Compass*" sulla valutazione della recidiva e quella italiana sulla mobilità degli insegnanti, l'autrice giunge a sostenere l'importanza del diritto "*ad una decisione umana*". Tale concetto, pur distinto, è strettamente affine alla nostra elaborazione di un "*diritto fondamentale all'interazione umana*". L'analisi dell'autrice, che evidenzia come un intervento umano sia necessario per "*validare, controllare ed anche smentire la decisione assunta autonomamente ed automaticamente dalla tecnologia*", fornisce un fondamento empirico alla nostra tesi, secondo cui la mera *human oversight* prevista dall'*AI Act* è insufficiente e va integrata da un diritto sostanziale a un rapporto dialogico con un altro essere umano.

[60] La distinzione è cruciale in O. POLLICINO, *Regolazione e innovazione tecnologica*, cit. 157, che contrappone la stagione dell'"*automazione algoritmica*", caratterizzata dall'esecuzione meccanica di istruzioni, a quella dell'"*autonomia*" dell'IA, in grado di "*prendere decisioni*" e porre sfide qualitativamente diverse per lo Stato di diritto.

[61] Art. 7 d.d.l. A.S. n. 1146-B.

[62] Art. 11 d.d.l. A.S. n. 1146-B.

[63] Art. 15 d.d.l. A.S. n. 1146-B.

[64] Artt. 7, 11, 13, 14, 15, d.d.l. A.S. n. 1146-B.

[65] Sul punto, si veda F. DONATI, *Intelligenza Artificiale e giustizia*, cit., 429, il quale esclude categoricamente la possibilità che un sistema di IA possa sostituirsi al giudice, argomentando l'incompatibilità di tale ipotesi con gli artt. 25, 101, 102 e 111 della Costituzione, che delineano un modello di giurisdizione fondato sulla figura del giudice-persona. C. SAGONE, *Efficientamento della giustizia*, cit., 312 ss., che esclude l'uso sostitutivo dell'IA nell'esercizio della funzione giurisdizionale in base agli artt. 24, 101 e 111 Cost.

che sono spesso i principali destinatari di tali servizi, sono i più penalizzati dalla potenziale assenza di un interlocutore umano capace di comprendere le loro specifiche esigenze.

L'assenza di interazione umana, dunque, non è una mera questione di bilanciamento tra "*efficienza*" e "*umanità*", ma diventa un "*ostacolo di ordine… sociale, che, limitando di fatto la libertà e l'eguaglianza dei cittadini, impedisce il pieno sviluppo della persona umana*", in violazione diretta del mandato costituzionale di cui all'art. 3, comma 2, della Costituzione[66].

Il "*diritto all'interazione umana*" non è, quindi, solo un'emanazione del principio personalista dell'art. 2 Cost., ma si rivela uno strumento fondamentale per l'attuazione del principio di uguaglianza sostanziale.

Garantire un canale di comunicazione umano non è una opzione, ma un dovere inderogabile della Repubblica per rimuovere le nuove barriere create dalla digitalizzazione e assicurare un accesso equo e non discriminatorio ai diritti, specialmente per le fasce più deboli della popolazione.

6. Le garanzie procedimentali alla prova dell'opacità algoritmica: dal diritto alla spiegazione alla tutela giurisdizionale effettiva.

La tutela dei diritti di fronte a decisioni automatizzate si articola su un complesso di garanzie procedurali e giurisdizionali che, tuttavia, si scontrano con il muro dell'opacità algoritmica. Il "*diritto alla spiegazione*", ovvero il diritto di ottenere "*informazioni significative sulla logica utilizzata*" sancito dall'art. 15, par. 1, lett. h), del GDPR[67], rappresenta il cardine di questo sistema di garanzie[68], specificato oggi, limitatamente alle decisioni basate sull'*output* di sistemi di IA classificati come "*ad alto rischio*", dall'art. 86 dell'*AI Act*, come "*come diritto di ottenere… spiegazioni chiare e significative sul ruolo del sistema di IA nella procedura decisionale e sui principali elementi della decisione adottata*".

Si tratta di una disposizione che, agendo come *lex specialis*, cristallizza un diritto esplicito e più definito, che affianca[69] e, in definitiva, rafforza il diritto ad avere informazioni significative sulla "*logica*" del sistema, previsto dal GDPR, come ulteriore diritto, nei casi previsti, ad una spiegazione vera e propria vera e propria, focalizzata non più sul funzionamento interno dell'algoritmo, quanto sul suo contributo effettivo ("*ruolo*") e sui fattori determinanti ("*principali elementi*") che hanno condotto alla decisione.

La Corte di Giustizia dell'Unione Europea, nelle fondamentali sentenze *Schufa Holding* (causa C-634/21) e *Dun & Bradstreet* (causa C-203/22), ha significativamente e a sua volta fortificato tale diritto, per come in origine previsto dal GDPR, chiarendo che esso non si

[66] La nostra proposta di un costituzionalismo algoritmico, specialmente in relazione all'uso dell'IA nella pubblica amministrazione, è corroborata dal lavoro di G. FASANO, *Le 'informazioni sintetizzate' generate dai large language models*, cit., 107-132.sulla digitalizzazione della P.A. L'autore sostiene che le nuove tecnologie debbano essere orientate alla realizzazione dei compiti costituzionali, in particolare la rimozione degli ostacoli che impediscono "*il pieno sviluppo della persona umana*" ai sensi dell'art. 3, comma 2, Cost. Questa visione sposa pienamente la nostra tesi secondo cui il "*diritto all'interazione umana*" è strumento essenziale per l'uguaglianza sostanziale. L'ancoraggio dell'azione amministrativa digitale al "*diritto a una buona amministrazione*" (art. 41 CDFUE e art. 97 Cost.) e la centralità della persona sono elementi che rafforzano la necessità di un approccio preventivo e sostanziale, in opposizione a un modello regolatorio che, come da noi criticato, si limita a gestire il rischio *ex post*.

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

[68] F. DONATI, *Intelligenza Artificiale e giustizia*, cit., 425, evidenzia come le disposizioni del GDPR, in particolare gli artt. 15 e 22, costituiscano un presidio fondamentale per assicurare la trasparenza e la correttezza delle decisioni automatizzate al di fuori del perimetro del procedimento amministrativo.

[69] Art. 86, par. 3, *AI Act*.

esaurisce nella comunicazione di formule matematiche complesse, ma richiede una spiegazione intelligibile dei criteri e delle procedure che hanno condotto a una decisione specifica, al fine di consentire all'interessato di contestarla efficacemente[70].

Ciononostante, si argomenta che tale garanzia, per quanto necessaria e pur se rafforzata dall'art. 86, rimanga uno strumento *ex post* intrinsecamente insufficiente di fronte all'opacità strutturale dei sistemi di *machine learning* più complessi, la cui logica interna, basata su modelli statistici non lineari, spesso non è ricostruibile in termini causali-deterministici e pienamente intellegibili dall'uomo[71].

In questo contesto, la formulazione dell'art. 86 dell'*AI Act* riflette un pragmatico e consapevole cambio di paradigma. Il legislatore europeo sembra spostare l'onere della trasparenza dalla ricerca, spesso infruttuosa, di una spiegazione *interna* e completa del processo algoritmico, a una spiegazione *esterna* e funzionale della decisione. L'enfasi sul "*ruolo del sistema*" e sui "*principali elementi*" mira a fornire all'individuo le informazioni materialmente utili per comprendere la decisione che lo riguarda e, soprattutto, per poterla contestare efficacemente in giudizio. Questo spostamento potrebbe rappresentare il preludio ad una rinuncia ad affrontare il problema della *black box*, tramite una riformulazione del diritto alla spiegazione in termini giuridicamente (e umanamente) più efficaci e gestibili, ma ovviamente è troppo presto per vedere come sarà interpretata e attuata questa fondamentale disposizione.

Nell'ordinamento italiano, in assenza ancora di una disciplina legislativa organica, è stata la giurisprudenza amministrativa a farsi carico di elaborare un sistema di garanzie. A partire dalle prime e principali sentenze del Consiglio di Stato n. 2270/2019 e n. 8472/2019[72], si è andato consolidando un *corpus* di principi di "*legalità algoritmica*": il principio di "*conoscibilità*" [73] (inteso come trasparenza "*rinforzata*"), che impone la piena accessibilità all'algoritmo e alla sua logica[74]; il principio di "*non esclusività*", che richiede un intervento

[70] Cfr. Corte di Giustizia UE, 7 dicembre 2023, causa C-634/21, *Schufa Holding*; Corte di Giustizia UE, 27 febbraio 2025, causa C-203/22, *Dun & Bradstreet Austria*.

[71] Un esempio di questa logica procedurale si ritrova nel *Code of Practice - Transparency Chapter*, il quale istituisce un *Model Documentation Form*. Questo modulo mira a fornire ai fornitori a valle e alle autorità competenti informazioni dettagliate sul modello, inclusi i processi di addestramento e valutazione. Sebbene utile per la tracciabilità e la conformità formale, tale documentazione descrive il "*cosa*" e il "*come*" del modello, ma non è in grado di fornire una spiegazione causale e intelligibile del "*perché*" di una specifica decisione generata dalla *black box*, confermando la natura esterna e a posteriori di queste garanzie di trasparenza. Per una riflessione sul diritto alla spiegazione in relazione all'art. 24 Cost., si veda A. SIMONCINI, *Il linguaggio dell'Intelligenza Artificiale*, cit., 24 ss.

[72] Cfr., da ultimo, Consiglio di Stato sez. VI, 6 giugno 2025, n. 4929: "*In proposito, la giurisprudenza di questo Consiglio (cfr. ad es. Consiglio di Stato sez. VI, 13 dicembre 2019, n. 8472) e la successiva evoluzione normativa (cfr. ad es. art. 30 d.lgs. 36 del 2023) hanno evidenziato come l'utilizzo di sistemi di intelligenza artificiale (dagli algoritmi al c.d.* machine learning*, compreso ogni meccanismo rientrante nella attuale nozione di AI, su cui cfr. da ultimo le "*guidelines on the definition of an artificial intelligence system established by Regulation EU 2024/1689*", ex artt. 3 e 96 dello stesso regolamento), da intendersi quale modulo procedimentale per lo svolgimento dell'attività autoritativa in modalità più efficienti, si accompagna ad una serie di principi ermeneutici ed applicativi tesi a garantire l'operatività del sistema e la tutela dei diritti e degli interessi coinvolti*"; sulla scia di Consiglio di Stato, Sez. VI, 8 aprile 2019, n. 2270; Consiglio di Stato, Sez. VI, 13 dicembre 2019, n. 8472.

[73] Sull'analisi di questo *corpus* giurisprudenziale, si veda C. NAPOLI, *Algoritmi, intelligenza artificiale e formazione della volontà pubblica*, cit., 336 ss., la quale sottolinea come il Consiglio di Stato abbia ricondotto l'algoritmo alla categoria dell'"*atto amministrativo informatico*", sottoponendolo ai principi di conoscibilità, imputabilità e non esclusività della decisione.

[74] Per un'analisi approfondita di questo *corpus* giurisprudenziale e del principio di "*conoscibilità*" come declinazione rafforzata della trasparenza, si veda A. SIMONCINI, *Il linguaggio dell'Intelligenza Artificiale*, cit., 26

umano qualificato (di controllo o di verifica) nel processo decisionale; e il principio di "*non discriminazione*", che impone di verificare e correggere gli errori sistematici (i *bias*) presenti nei dati e nel modello[75].

Questi principi generali, codificati oggi nell'art. 30 del codice dei contratti pubblici[76], rappresentano un lodevole tentativo pretorio, prima, e normativo, poi, di positivizzare l'azione amministrativa algoritmica, ancorandola alle garanzie del giusto procedimento[77] e divenendo, quelli codicistici, parametri interpretativi anche in assenza di una disciplina organica, oltreché principi generali da attuare nello svolgimento di ogni funzione pubblica, influenzabile dall'uso di sistemi di IA o avente ad oggetto gli stessi.

Tuttavia, anche il sindacato giurisdizionale incontra limiti invalicabili. Il controllo del giudice amministrativo, pur estendendosi a un vaglio di "*ragionevolezza*" e di non manifesta illogicità dell'algoritmo, rimane un controllo estrinseco e formale.

Esso può verificare la correttezza degli *input* e la coerenza degli *output* rispetto alle regole predefinite, ma non può penetrare la *black box* per valutare la sostanza del "*ragionamento*" della macchina, specialmente quando questo investe valutazioni complesse riconducibili alla discrezionalità tecnica o, a maggior ragione, a quella amministrativa. Ciò pone un serio problema di effettività della tutela giurisdizionale, garantita come diritto inviolabile dagli artt. 24 e 113 della Costituzione[78].

Se il cittadino non può comprendere appieno le ragioni di una decisione che lo pregiudica, e il giudice non può sindacarle nella loro interezza, il diritto di difesa rischia di essere svuotato di contenuto. Per superare i limiti intrinseci di una tutela meramente *ex post*, si ritiene necessario un radicale cambio di prospettiva, abbracciando l'approccio del costituzionalismo algoritmico sostanziale[79].

Questo passaggio da un controllo esterno e successivo a un'integrazione interna e preventiva dei principi costituzionali rispecchia l'evoluzione stessa del costituzionalismo da "*moderno*" a "*contemporaneo*". Mentre il primo si concentrava su limiti formali e procedurali al potere (es. la separazione dei poteri), il costituzionalismo contemporaneo si caratterizza per

ss., il quale sottolinea come il giudice amministrativo abbia imposto la piena accessibilità all'algoritmo, inteso come il vero "*motore*" della decisione amministrativa.

[75] Cfr. F. DONATI, *Intelligenza Artificiale e giustizia*, cit., 424, che, analizzando la sentenza del Consiglio di Stato n. 2270 del 2019, sottolinea come l'algoritmo debba essere considerato un "*atto amministrativo informatico*", pienamente soggetto ai principi di pubblicità e trasparenza e sindacabile dal giudice amministrativo.

[76] Art. 30 D. Lgs. 31 marzo 2023, n. 36, recante "*Codice dei contratti pubblici in attuazione dell'articolo 1 della legge 21 giugno 2022, n. 78, recante delega al Governo in materia di contratti pubblici*".

[77] Per un'analisi approfondita dell'art. 30 del d.lgs. 36/2023, si veda D. U. GALETTA, *Digitalizzazione, Intelligenza artificiale e Pubbliche Amministrazioni*, cit., X, la quale definisce la norma come "*la prima previsione normativa che identifica i principi da rispettare in caso di utilizzo di procedure automatizzate*", individuando i quattro principi fondamentali della "*conoscibilità, comprensibilità, non esclusività e non discriminazione*".

[78] Questa argomentazione è al centro della tesi di A. SIMONCINI, *Il linguaggio dell'Intelligenza Artificiale*, cit. 34 ss., secondo cui un linguaggio tecnologico incomprensibile rende la decisione non giustiziabile, frustrando il diritto inviolabile alla difesa sancito dall'art. 24 Cost. e fondando un vero e proprio "*diritto costituzionale ad una tecnologia "ragionevole"*". L'esigenza di un "*costituzionalismo algoritmico sostanziale*", che integri i principi costituzionali direttamente nella progettazione (*by design*), è fortemente sostenuta da una parte significativa della dottrina. C. CASONATO, *Costituzione e intelligenza artificiale*, cit., 13, afferma che "*è essenziale, infatti, che il diritto non insegua le applicazioni della Al, ma che intervenga a monte, ponendo principi e regole by design - per così dire*". Similmente, T. GROPPI, *Alle frontiere dello Stato costituzionale*, cit., 683, invoca la necessità che i principi della democrazia costituzionale diventino "*senso comune, di permeare di sé la nostra epoca e chi in essa abita, compresi coloro che delle innovazioni… sono i protagonisti, gli scienziati*".

[79] o "*Constitutional by Design*", elaborato in dottrina da G. DE MINICO, *Towards an "Algorithm Constitutional by Design"*, cit., 381 ss.

la sua dimensione "*sostanziale*": non si limita a chiedere *come* il potere viene esercitato, ma impone *vincoli di contenuto*, radicati in valori come la dignità, l'uguaglianza e la solidarietà.

Il costituzionalismo algoritmico sostanziale, dunque, non fa altro che applicare questa logica al mondo della tecnologia, chiedendo che l'algoritmo non sia solo formalmente corretto, ma intrinsecamente "*giusto*" e rispettoso dei valori fondamentali.

Questo paradigma non si limita a verificare la conformità di un sistema algoritmico a regole esterne, ma impone di integrare i principi costituzionali - dignità, uguaglianza, non discriminazione - direttamente nella fase di progettazione (*by design*) e di funzionamento predefinito (*by default*) dell'algoritmo[80].

Ciò implica un obbligo giuridico ancorato nei principi supremi dell'ordinamento democratico per i progettisti e i fornitori di creare sistemi che siano, per loro stessa architettura, trasparenti, equi, spiegabili e contestabili.

Il modello regolatorio attuale concepisce il diritto come un limite esterno alla tecnica: il diritto interviene per correggere, sanzionare o imporre obblighi di trasparenza a un artefatto tecnologico già esistente e spesso intrinsecamente opaco. Questo approccio si rivela inefficace perché la logica interna della tecnologia rimane impermeabile al controllo giuridico tradizionale. Il diritto può così solo osservare gli *input* e gli *output*, ma il processo che li connette rimane una scatola nera.

L'approccio sostanziale ribalta questa prospettiva: non è più il diritto che deve adattarsi faticosamente e farraginosamente alla tecnica, ma è la tecnica che deve essere progettata e sviluppata per essere costituzionalmente compatibile[81].

Questo approccio trasforma e ri-potenzia i principi fondamentali costituzionali da vincoli esterni a requisiti di progettazione interni, rendendo la conformità costituzionale una precondizione per la legalità stessa della tecnologia.

---

[80] Anche il *Code of Practice - Safety and Security Chapter*, pur essendo il più avanzato, si ferma a un passo da questo paradigma. Esso richiede ai fornitori di modelli a rischio sistemico di documentare le "*valutazioni del modello*" e le "*strategie*" adottate, nonché le "*mitigazioni di sicurezza*" implementate. Questo rappresenta un importante obbligo di documentazione del processo, ma non un obbligo di progettare sistemi la cui architettura interna sia intrinsecamente spiegabile e costituzionalmente compatibile. Parte della dottrina, pur criticando aspramente l'opacità, si concentra sulla necessità di assicurare garanzie procedurali che rendano la decisione algoritmica "*visibile*", "*intelligibile*" e, quindi, "*giustiziabile*". In questa prospettiva, la sfida è assicurare la trasparenza come "*misura di effettività dell'equazione: conoscibilità dell'atto/sua sfidabilità in giudizio*", proponendo un modello di "*Algorithm Constitutional by Design*" fondato su tali garanzie. Cfr. G. DE MINICO, *Towards an "Algorithm Constitutional by Design"*, cit., 391; G. DE MINICO, *Giustizia e intelligenza artificiale*, cit., 89. L'esigenza di integrare le garanzie sin dalla fase di progettazione è un principio cardine del c.d. "trustworthiness & security", secondo cui tutte le iniziative "*dovrebbero andare oltre il semplice rispetto del quadro normativo in materia di protezione dei dati personali, privacy e sicurezza informatica, integrando tali elementi già nella fase di progettazione*", cfr. D. U. GALETTA, *Digitalizzazione, Intelligenza artificiale e Pubbliche Amministrazioni*, cit., VII. Tale approccio è coerente con le linee guida etiche della Commissione europea, che, come ricorda G. MAIRA, *Intelligenza umana e intelligenza artificiale*, cit., XVII, puntano "*principalmente sulla centralità dell'essere umano: prima degli algoritmi devono venire la dignità e la libertà dell'uomo*".

[81] Questo rovesciamento di prospettiva trova un profondo fondamento teorico nel lavoro di Mireille Hildebrandt, la quale ha indagato come l'infrastruttura tecnologica della "*data-driven agency*" stia alterando la natura stessa del diritto. Per Hildebrandt, il diritto moderno è un prodotto della cultura della stampa, basato sulla testualità, l'interpretazione e la contestabilità. L'avvento di un ambiente "*onlife*" governato da sistemi che agiscono sulla base di correlazioni statistiche e non di norme interpretabili minaccia le "*condizioni di esistenza*" dello Stato di diritto. Cfr. M. HILDEBRANDT, *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology*, Edward Elgar Publishing, 2016. La proposta di un costituzionalismo algoritmico sostanziale risponde a questa sfida, tentando di re-inscrivere la logica della norma giuridica e della garanzia costituzionale all'interno della nuova infrastruttura computazionale.

Non si tratta di una mera proposta di *soft law* o di un principio etico, ma di un principio giuridico-costituzionale cogente che fondi una nuova forma di legalità sostanziale per la tecnologia.

Occorrerebbe prendersi atto dell'opacità dei sistemi di IA i cui modelli e processi decisionali non sono suscettibili di essere riesaminati in modo significativo o compresi in termini umanamente intelligibili, a causa della loro complessità tecnica, e che in essi il contributo specifico degli *input* a un dato *output* non può essere pienamente tracciato e spiegato e trarne le dovute conseguenze.

Questa categoria di sistemi impone obblighi più stringenti, come la spiegabilità per progettazione (*explainability by design*) e valutazioni di divieti d'uso in contesti sensibili.

Per dare sostanza al paradigma del costituzionalismo algoritmico, essendo oggi impensabile una revisione dell'*AI Act*, è cruciale la tesi di chi[82] valorizza il contratto pubblico come strumento strategico e primario di regolazione. In questa prospettiva, l'acquisto di sistemi di intelligenza artificiale da parte della pubblica amministrazione diventa il luogo preventivo in cui i principi costituzionali possono già sostanziarsi ed essere tradotti in requisiti concreti e l'esecuzione del contratto il luogo *in itinere*, in cui può essere costantemente monitorata la conformità costituzionale dell'uso e dei risultati prodotti dai sistemi di IA.

È attraverso le clausole contrattuali e il governo della fase esecutiva, come sottolinea detta dottrina, che si può imporre il rispetto dei pilastri della legalità algoritmica: la trasparenza e l'intellegibilità dei processi, la non esclusività della decisione automatizzata e le garanzie contro la discriminazione.

Il contratto cessa così di essere un mero atto di acquisto per diventare lo strumento che struttura l'esercizio stesso delle funzioni pubbliche, creando un'intersezione inedita in cui le pratiche contrattuali sono al contempo strumento e oggetto di regolamentazione.Ci pare questa la più concreta via per garantire una tutela giurisdizionale che sia *effettiva*, come imposto dagli artt. 24 e 113 della Costituzione, e non meramente formale.

7. Traiettorie interpretative e attuative del "*costituzionalismo algoritmico sostanziale*" a legislazione invariata.

Per superare le criticità intrinseche all'approccio basato sul rischio e dare concretezza al paradigma del costituzionalismo algoritmico sostanziale, è possibile e necessario agire sul piano interpretativo e attuativo.

Un ruolo cruciale in questo percorso di riorientamento costituzionale è affidato alle future Autorità Nazionali Competenti (NCA). L'*AI Act* (art. 70) impone agli Stati membri di designare almeno un'Autorità di notifica e un'Autorità di vigilanza del mercato; il disegno di legge italiano è lo strumento normativo chiamato a individuare specificamente tali organismi nell'ordinamento interno. Queste Autorità, insieme alla giurisprudenza, detengono le chiavi per orientare l'applicazione della normativa in senso autenticamente antropocentrico.

Tuttavia, un'analisi pragmatica impone di riconoscere che non tutte le traiettorie che proponiamo hanno la medesima probabilità di successo nel breve periodo. È opportuno, quindi, adottare un approccio graduale e strategico, distinguendo tra obiettivi immediatamente perseguibili e traguardi a medio termine che richiedono un'azione coraggiosa da parte delle Autorità.

La via più solida e immediatamente percorribile è quella del rafforzamento della tutela giurisdizionale. Questa traiettoria non rappresenta una forzatura del dato normativo, ma si

[82] G. F. LICATA, *Intelligenza artificiale e contratti pubblici*, cit.

pone in linea di continuità con un'evoluzione già in atto, trainata dalla giurisprudenza sopra citata della Corte di Giustizia dell'UE in materia di GDPR.

I requisiti di trasparenza (art. 13 *AI Act*) e il diritto alla spiegazione (art. 86 dell'*AI Act*), per i sistemi "*ad alto rischio*" devono essere interpretati rigorosamente alla luce del diritto a un ricorso giurisdizionale effettivo e del diritto di difesa (art. 47 CDFUE e artt. 24 e 113 Cost.). In questo senso, gli obblighi di trasparenza previsti dal *Code of Practice*, come quello di fornire un riassunto pubblico delle *policy* sul *copyright* e dei contenuti usati per l'addestramento, o di documentare dettagliatamente le valutazioni di sicurezza, devono essere considerati non come meri adempimenti formali, ma come strumenti abilitanti per l'esercizio effettivo del diritto di difesa. Una spiegazione può considerarsi "*chiara e significativa*" ai sensi dell'art. 86 solo se, attingendo a tale documentazione, permette di ricostruire la catena logica che ha portato alla decisione, rendendola pienamente contestabile in giudizio.

Un sistema di tal fatta può ritenersi conforme solo se la sua spiegabilità consente la piena difesa in giudizio, fornendo la *ratio decidendi* (la logica giuridico-fattuale della decisione specifica), non una generica "*descrizione tecnica*" del modello di IA.

In questa prospettiva, la portata dell'art. 86 dell'*AI Act* è particolarmente incisiva. Esso impone una spiegazione che verta sugli specifici dati di *input* che hanno avuto un peso preponderante e sulla ricostruzione della catena di inferenze applicata al caso concreto, partendo dalle "*massime operative*" generali del sistema per arrivare alla conclusione particolare. L'obiettivo è quello di garantire che la tutela giurisdizionale non sia vanificata dall'opacità, assicurando che ogni decisione, anche se assistita da IA, rimanga pienamente sindacabile nella sua logica e nelle sue fondamenta fattuali. Il passaggio dalla spiegazione della "*logica del sistema*" alla spiegazione della "*decisione*" non deve comunque legittimare un implicito abbandono del problema della *black box*, a favore soltanto della spiegazione dell'*output* decisionale in termini umanamente intellegibili.

Di fronte all'asimmetria informativa strutturale generata dalla *black box*, inoltre, l'evoluzione giurisprudenziale potrebbe orientarsi verso un alleggerimento dell'onere della prova per il danneggiato, richiedendo all'utilizzatore del sistema opaco di dimostrarne la correttezza e la non discriminatorietà in caso di presunta violazione di diritti e ciò anche per il principio di vicinanza della prova.

La trasformazione della sorveglianza umana da adempimento formale a presidio di giustizia sostanziale è un obiettivo plausibile, ma giuridicamente più complesso. Sebbene il disegno di legge italiano ne enfatizzi il carattere di principio[83], la clausola che vieta l'introduzione di "*nuovi obblighi*"[84] rispetto all'*AI Act* potrà rappresentare un ostacolo significativo.

Questo percorso richiede un'azione politicamente orientata e coraggiosa da parte delle Autorità nazionali competenti, che dovrebbero adottare linee guida rigorose che interpretino l'*human oversight* (Art. 14 *AI Act*), quantomeno come obbligo di riesame critico e sostanziale, non come mera validazione automatica (*rubber-stamping*). Questo implica garantire che l'operatore umano disponga dell'autorità, della competenza e del tempo necessari per disattendere l'*output* algoritmico non solo per errori tecnici, ma per ragioni di equità e giustizia nel caso concreto.

Parallelamente, sfruttando le competenze nazionali sul procedimento amministrativo, è fondamentale rafforzare il diritto al contraddittorio umano, assicurando al cittadino, in particolare per le valutazioni discrezionali complesse che coinvolgono il bilanciamento di

[83] Art. 3, comma 3, d.d.l. A.S. n. 1146-B.
[84] Art. 3, comma 5, d.d.l. A.S. n. 1146-B.

diritti fondamentali, la possibilità effettiva di interloquire direttamente con un funzionario responsabile qualora contesti una decisione assistita dall'IA, dando così sostanza all'interazione umana.

Infine, è necessario valorizzare, come già anticipato, lo strumento alternativo, trasversale e di immediata efficacia del contratto pubblico. Le pubbliche amministrazioni, in qualità di grandi acquirenti di tecnologia, possono agire come regolatori di fatto del mercato. Inserendo nei bandi di gara clausole specifiche e vincolanti, possono trasformare i principi del costituzionalismo algoritmico in requisiti di mercato. Tali clausole possono imporre ai fornitori, come condizione per l'aggiudicazione, obblighi di *explainability by design*, *audit* di parte terza sulla non discriminatorietà, architetture intrinsecamente trasparenti e meccanismi rafforzati di interazione umana. In questo modo, il contratto pubblico cessa di essere un mero atto di acquisto per diventare la leva più attuale per orientare l'innovazione, spostando il baricentro del controllo dalla gestione del danno *ex post* alla garanzia della compatibilità costituzionale *ex ante*.

8. Conclusione: per un modello di sviluppo tecnologico radicato nei diritti umani.

L'analisi condotta ha evidenziato le profonde criticità del quadro normativo che si sta delineando in materia di Intelligenza Artificiale.

L'approccio funzionalista nella definizione e quello basato sulla gestione del rischio si rivelano strumenti inadeguati, che conducono a una tutela dei diritti fondamentali meramente formale e correttiva[85]. Tale impostazione, anziché governare la trasformazione tecnologica, rischia di subirla, normalizzando l'opacità decisionale e riducendo i diritti fondamentali a variabili negoziabili in un calcolo di convenienza. Sembra necessario, pertanto, un radicale cambio di paradigma, che abbandoni la logica della gestione del rischio per abbracciare un modello di regolazione fondato su un costituzionalismo algoritmico sostanziale. In questa prospettiva, i diritti fondamentali non possono essere meri "*rischi*" da mitigare, ma devono rappresentare il fine ultimo e il limite invalicabile di ogni sviluppo tecnologico[86].

In definitiva, si tratta di riaffermare la funzione classica del costituzionalismo: garantire che nessuna forma di potere - politico, economico o, oggi, tecnologico - possa operare in uno spazio privo di diritto (*legibus solutus*). L'opacità della *black box* algoritmica rappresenta la frontiera contemporanea di questo spazio. Lasciare che le decisioni che incidono sui diritti fondamentali vengano prese all'interno di questa scatola nera equivarrebbe a un ritorno a forme di potere assoluto e arbitrario che le costituzioni moderne sono nate per contrastare.

Un approccio autenticamente preventivo dovrebbe fondarsi su due pilastri.

Il primo è il riconoscimento del "*diritto all'interazione umana*" come diritto fondamentale, presidio indispensabile della dignità e dell'uguaglianza sostanziale, in un mondo sempre più automatizzato. Esso garantisce in radice che la tecnologia rimanga uno strumento al servizio della persona e non un sostituto delle relazioni umane essenziali per decisioni giuste ed eque.

---

[85] In una prospettiva parzialmente dialettica rispetto alla critica di un vuoto normativo, si può osservare come l'uso della *soft law* sia una scelta deliberata del legislatore europeo per regolare settori in rapida evoluzione. P. INTURRI, *Intelligenza artificiale e soft law,* cit., 347, 350, chiarisce che la *soft law* può assumere una funzione di para-*law*, ponendosi come "*alternativa a norme di hard law*". In questa visione, strumenti come i codici di condotta non sono necessariamente una delega in bianco, ma un meccanismo di regolazione flessibile che evita di cristallizzare in legge norme che diventerebbero rapidamente obsolete, permettendo un adattamento più rapido alle innovazioni tecnologiche.

[86] Cfr. F. DONATI, *La protezione dei diritti fondamentali*, cit., 1; O. POLLICINO, *Regolazione e innovazione tecnologica*, cit., 119 ss.

Questo diritto va ben oltre la mera "*sorveglianza umana*" (*human oversight*) prevista dall'art. 14 dell'*AI Act*, che si configura come un controllo funzionale, spesso formale e successivo, sull'*output* della macchina. Il diritto all'interazione umana è, invece, un diritto sostanziale a un rapporto dialogico, empatico e comprensivo con un altro essere umano durante ogni procedimento decisionale che incide sulla sfera giuridica di una persona. Questo diritto opera come una "*riserva di umanità*" rinforzata, un limite costituzionale invalicabile alla delega di funzioni pubbliche a sistemi automatizzati, specialmente quando esse implicano l'esercizio di discrezionalità. Esso garantisce che la tecnologia rimanga uno strumento al servizio della persona, veicolando valori non computabili come l'empatia, la comprensione del contesto e il riconoscimento della dignità.

Il secondo e, più strutturale, pilastro è l'adozione del principio del costituzionalismo algoritmico sostanziale, che impone di inscrivere le garanzie costituzionali - trasparenza, equità, non discriminazione, giustiziabilità -nell'architettura stessa dei sistemi algoritmici. Solo obbligando la tecnica a essere progettata in conformità con il diritto, e non viceversa, si può superare il problema della *black box* e assicurare una tutela che sia realmente *ex ante* ed effettiva.

Questo principio trasforma le garanzie costituzionali - trasparenza, equità, non discriminazione, spiegabilità - da vincoli esterni a requisiti di progettazione interni. Esso stabilisce un obbligo giuridico cogente per sviluppatori e fornitori di integrare tali garanzie direttamente nell'architettura tecnica dei sistemi di IA. È il rimedio tecnico-giuridico diretto al problema dell'opacità.

Invece di tentare di decifrare una *black box a posteriori*, si dovrebbe prevedere che, per essere legalmente impiegato, quanto meno in settori ad alto impatto e in contesti sensibili, un sistema debba essere progettato in modo tale che le sue decisioni siano tracciabili, verificabili e spiegabili in termini umanamente intelligibili.

L'onere della conformità si dovrebbe spostare così dalla documentazione *ex post* (come le valutazioni d'impatto) alla validazione architettonica *ex ante*.

I due pilastri sono profondamente interconnessi e si rafforzano a vicenda. Il secondo mira in radice a rendere il "*ragionamento*" della macchina il più trasparente possibile, fornendo le fondamenta tecniche per la responsabilità. Tuttavia, anche un sistema perfettamente spiegabile può produrre una decisione formalmente corretta, ma sostanzialmente ingiusta, perché incapace di cogliere il contesto umano non quantificabile. È qui che interviene il "*diritto all'interazione umana*", fornendo il ricorso necessario a un giudizio umano finale, capace di empatia e discrezionalità.

Insieme, questi due pilastri creano un sistema di tutela completo: il pilastro tecnico garantisce che la macchina sia "*responsabile*", mentre il pilastro umano assicura che il processo rimanga "*giusto*" e "*umano*".

Un simile approccio può assicurare che lo sviluppo dell'IA sia orientato al benessere collettivo, al pieno sviluppo della persona umana e al rafforzamento delle istituzioni democratiche.

Questa è la base per una vera "*sovranità algoritmica*" europea: non una sovranità definita dalla mera competizione tecnologica o dalla capacità di imporre *standard* di mercato, ma dalla capacità di plasmare la tecnologia in conformità con i propri valori costituzionali più profondi.

La scelta di un paradigma fondato sul "*diritto all'interazione umana*" e sul "*costituzionalismo algoritmico sostanziale*" non è, dunque, una mera opzione tecnico-giuridica, ma una decisione politica e di civiltà che può definire il ruolo dell'Europa nel mondo digitale del XXI secolo.

Si tratta, in definitiva, di un modello di sviluppo umano-centrico che non affidi del tutto la *governance* dell'IA alle forze del mercato, rischiando di intensificare le disuguaglianze e di depotenziare i diritti fondamentali e che riaffermi con forza la primazia della persona e dei suoi diritti inviolabili. È un modello in cui l'innovazione non è fine a se stessa, ma è strumento per la piena realizzazione della dignità umana, l'attuazione dell'uguaglianza sostanziale e il rafforzamento delle istituzioni democratiche.

Questa è, secondo noi, la sfida e, al tempo stesso, la più autentica missione del costituzionalismo nell'era dell'intelligenza artificiale.