

Senato della Repubblica
Commissione straordinaria per la tutela e la promozione dei diritti umani
Audizione della prof.ssa Ginevra Cerrina Feroni
(9 maggio 2024)

Ringrazio la Commissione e la sua Presidente, Sen. Stefania Pucciarelli, per l'invito a partecipare a questo confronto, che considero quantomai importante e, direi anche, urgente. Tra gli aspetti meno discussi dell'I.A. vi è, infatti, quello relativo alle sue implicazioni – che si prefigurano strutturali e profonde - sui diritti umani, nonostante l'evidente e trasversale incidenza di questa tecnologia “disruptive” sulle libertà e sui diritti fondamentali.

La peculiarità caratterizzante l'I.A. è, infatti, come ha osservato il filosofo francese Eric Sadin, il suo essere una tecnica non tanto e non solo protesica - volta cioè a compensare carenze dell'uomo - quanto piuttosto mimetica, capace cioè di emulare la razionalità umana e di apprendere autonomamente. Questi elementi rendono l'I.A. assolutamente innovativa e non realmente paragonabile alle altre tecnologie, anche per l'imprevedibilità che può connotarne talora i processi nonché per l'autonomia, rispetto alla decisione umana, che spesso può caratterizzarne gli esiti.

La capillare diffusione e la trasversalità dell'I.A. a ogni settore della vita privata e pubblica, molto più significative di quelle delle altre tecnologie, evidenziano inoltre l'esigenza di definire un limite di ammissibilità etica, politica e, quindi, anche giuridica degli usi di I.A. Ciò dal momento che, come osservava già Rodotà venticinque anni orsono, non tutto ciò che è tecnicamente possibile è per ciò stesso eticamente ammissibile, giuridicamente lecito, socialmente accettabile.

La continua espansione ed evoluzione dell'I.A. impone dunque di tracciare (e questo è il massimo compito della politica) un **limite di sostenibilità**, delle colonne d'Ercole da non varcare perché il progresso non divenga, paradossalmente, socialmente regressivo.

Si pensi, a titolo meramente esemplificativo, ai progetti di utilizzo dell'I.A. in campo neuroscientifico, con la realizzazione di *decoder* “semantici” dell'attività neurale, combinando scansione cerebrale e database di modelli linguistici, come quelli usati da Chat Gpt. A gennaio è stato applicato, per la prima volta, a un paziente tetraplegico, un dispositivo in grado di decodificare i segnali neurali per far eseguire a un robot ciò che i suoi arti non possono fare.

Si tratta di un'innovazione potenzialmente rivoluzionaria, capace di apportare benefici senza precedenti per la cura di stati neurodegenerativi e, per ciò, meritevole di sviluppo, purché tuttavia non si giunga alla trasparenza del pensiero: da sempre il sogno delle dittature. La possibilità di traduzione del pensiero in impulsi algoritmici è, infatti, una conquista preziosa a condizione che non venga utilizzata per leggerlo, rendendo dunque accessibile anche quel foro interno la cui riservatezza è presupposto necessario per la libertà di coscienza. L'applicazione dell'intelligenza artificiale in campo neuroscientifico e, soprattutto i sistemi di *brain reading*, idonei almeno potenzialmente a decodificare il pensiero, devono quindi sempre garantire, come primo dei diritti umani, la privacy mentale, condizione ineludibile di autodeterminazione, presupposto intangibile di libertà. Varcata la soglia della lettura del pensiero, la deriva da impedire è rendere la persona un archivio liberamente accessibile, le cui idee siano messe a nudo senza più alcuno spazio per la libertà, anzitutto di determinazione. Si tratta di uno dei tanti possibili esempi di come l'I.A. esiga una regolazione attenta, in primo luogo, a indirizzarne lo sviluppo al servizio dell'uomo e non viceversa, per la promozione e non certo per la violazione dei diritti umani.

Un altro aspetto meritevole di considerazione riguarda i **bias** (distorsioni, pregiudizi algoritmici) suscettibili di caratterizzare il processo decisionale algoritmico, con esiti spesso discriminatori. L'algoritmo non è, infatti, neutro ma riflette, persino amplificandole, le precomprensioni di chi lo progetta e l'eventuale incompletezza dei dati con i quali viene addestrato. Ciò determina spesso esiti discriminatori più ancora di quelli che, paradossalmente, la fredda razionalità della macchina avrebbe dovuto evitare. Sono numerosi i casi di arresti di persone del tutto estranee ai fatti per l'errore dell'algoritmo di riconoscimento facciale utilizzato in alcuni Stati americani, addestrato su una serie non sufficientemente rappresentativa dei tratti somatici di tutte le etnie.

La sub-rappresentatività dello spettro di dati su cui si addestrano gli algoritmi si è rivelata essere profondamente distorsiva soprattutto se utilizzata in sede giudiziaria. Si pensi all'algoritmo di prognosi di recidiva penale Compas, utilizzato da alcune corti americane, incline ad assegnare un tasso di propensione alla recidiva maggiore agli afroamericani, perché allenato sui dati statistici relativi alla composizione etnica della popolazione penitenziaria statunitense (per gran parte costituita da quanti, privi dei mezzi per accedere alla cauzione, non possono che scontare la pena in carcere). L'algoritmo inferiva, insomma, dal dato effettuale, inevitabilmente parziale, dell'appartenenza etnica dei detenuti un nesso causale in realtà inesistente dal punto di vista criminologico viziando, dunque, con un pregiudizio discriminatorio, il giudizio sul rischio di recidiva. E se sul giudiziario la discriminazione algoritmica ha effetti più significativi, essa

comunque può avere un'incidenza sui diritti umani, sul principio di eguaglianza e sulla stessa tutela della dignità personale non meno rilevante in ogni altro settore in cui si manifesti. Vi è, ad esempio, anche questa preoccupazione alla base del divieto di social scoring introdotto dall'Artificial Intelligence Act, con il rischio di derive antidemocratiche come quelle della “cittadinanza a punti” del sistema cinese.

Un altro punto di incidenza significativo e “di sistema” dell'I.A. sui diritti umani riguarda il suo utilizzo nel settore della *law enforcement*. Il ricorso all'I.A. in ambito investigativo necessita, infatti, cautele tali da scongiurare il rischio che si deleghino all'algoritmo attività potenzialmente incidenti sulla libertà personale e di sorveglianza massiva. Ciò che si teme non è tanto e non è solo il “pendio scivoloso”, quanto la tendenza all'acritica accettazione sociale di una progressiva limitazione della libertà. Vi è questa preoccupazione, ad esempio, alla base dei divieti e dei limiti posti, con indubbia lungimiranza, ai sistemi di polizia predittiva e categorizzazione biometrica **dall'Artificial Intelligence Act.**

E per comprendere il grado di incidenza dell'I.A. sui diritti umani è utile proprio **partire da questo testo**, particolarmente significativo in quanto prima disciplina al mondo, di taglio generale, dell'I.A., considerando che quella americana è un mero Executive Order rivolto, come tale, alle sole agenzie federali e dal contenuto alquanto limitato. Ed è significativo che la “primazia”, cronologica ma anche assiologica, nella regolazione di questa straordinaria tecnologia spetti all'Europa. A ciascuna delle norme dell'AI Act è, infatti, sottesa una linea chiara di politica del diritto, che esprime la stessa identità dell'Unione Europea come “comunità di diritto”. L'aspirazione a rendere l'I.A. “trustworthy”, affidabile, racconta molto, infatti, dell'idea europea di innovazione come tale solo se sostenibile dal punto di vista democratico e dello Stato di diritto, secondo una valutazione che si fonda in primo luogo sulla garanzia dei diritti e delle libertà fondamentali, sanciti dalla Carta di Nizza.

L'art. 1, oltre che il C 1, individua infatti nella tutela dei diritti fondamentali uno degli scopi principali della regolazione dell'I.A. e in tutto il corpo normativo essa costituisce tanto obiettivo di tutela quanto parametro di ammissibilità- oltre che standard regolatorio- dei sistemi algoritmici. L'incidenza sui diritti fondamentali è, infatti, criterio di valutazione del rischio connesso al sistema di I.A. che ne orienta la collocazione all'interno della complessa tassonomia stilata dal regolamento e, allo stesso tempo, oggetto specifico della valutazione d'impatto necessaria per il ricorso a determinati sistemi di I.A. ad alto rischio.

Il C 48 chiarisce, infatti, che *“La portata dell'impatto negativo del sistema di IA sui diritti fondamentali protetti dalla Carta è di particolare rilevanza ai fini della classificazione di un sistema di IA tra quelli ad alto rischio. Tali diritti comprendono il diritto alla dignità umana, il rispetto della vita privata e della vita familiare, la protezione dei dati personali, la libertà di espressione e di informazione, la libertà di riunione e di associazione e la non discriminazione, il diritto all'istruzione, la protezione dei consumatori, i diritti dei lavoratori, i diritti delle persone con disabilità, l'uguaglianza di genere, i diritti di proprietà intellettuale, il diritto a un ricorso effettivo e a un giudice imparziale, i diritti della difesa e la presunzione di innocenza e il diritto a una buona amministrazione”* oltre che le specifiche garanzie accordate ai minori.

E' inoltre significativo che: la definizione di incidente e rischio sistemico si fondi sull'impatto sui diritti fondamentali; che esso sia oggetto della sorveglianza umana imposta dall'art. 14; quella relativa ai diritti umani sia una competenza specifica di cui deve disporre il personale dell'autorità di notifica (art. 28, p.7) e in genere delle autorità competenti (art. 70, p.3); l'incidenza sui diritti fondamentali sia parametro di valutazione per le sperimentazioni normative (art. 57, p.11); le autorità con compiti di tutela di tali diritti siano assegnatarie di specifici poteri (art. 77) e destinatarie di un puntuale obbligo di consultazione (art. 82); l'incidenza sui diritti fondamentali radichi uno specifico diritto alla spiegazione del processo decisionale di un sistema di I.A. ad alto rischio (art. 86).

Il C 110, inoltre, include tra i rischi sistemici suscettibili di derivare dall'I.A. con fini generali le *“modalità con cui i modelli possono dar luogo a dannosi pregiudizi e discriminazioni con rischi per gli individui, le comunità o le società; l'agevolazione della disinformazione o la violazione della vita privata con minacce ai valori democratici e ai diritti umani”*.

Un'attenzione peculiare è poi riservata, dal C 60, ai sistemi ad alto rischio utilizzati nel settore della migrazione, dell'asilo e del controllo delle frontiere in ragione della particolare vulnerabilità degli interessati, tale per cui *“l'accuratezza, la natura non discriminatoria e la trasparenza dei sistemi di IA utilizzati in tali contesti sono pertanto particolarmente importanti per garantire il rispetto dei diritti fondamentali delle persone interessate, in particolare i loro diritti alla libera circolazione, alla non discriminazione, alla protezione della vita privata e dei dati personali, alla protezione internazionale e alla buona amministrazione”*. La vigilanza sull'uso dell'i. a. in tali contesti, per attività di contrasto, oltre che per fini di “giustizia e democrazia” è, peraltro, specificamente riservata (art. 74, p.8) alle autorità di protezione dati, nella consapevolezza, evidentemente, dell'esperienza maturata nell'applicazione di quello che è il fulcro dell'I.A.: il processo decisionale automatizzato fondato su dati personali. Rispetto ad esso, tanto il GDPR quanto la direttiva 2016/680 sul trattamento dei dati personali nel settore della giustizia penale e della polizia assicurano, infatti, alcune

garanzie essenziali: il principio di conoscibilità (che esclude la legittimità di algoritmi *black-box* riconoscendo il diritto di ricevere informazioni significative sulla logica utilizzata), quello di non esclusività della decisione algoritmica che impone un intervento umano capace di controllare, validare o smentire la decisione automatizzata, il divieto di discriminazione algoritmica, un generale principio di trasparenza che impone precisi obblighi informativi nei confronti dell'utente, un criterio di qualità ed esattezza dei dati da utilizzare, particolarmente rilevante per evitare i *bias* propri di un addestramento dell'algoritmo sulla base di informazioni inesatte o non sufficientemente rappresentative. Le garanzie particolari accordate nel trattamento dei dati dei minori si sono, inoltre, rivelate determinanti nell'assicurare il doveroso controllo sull'accesso degli infraquattordicenni ad alcuni dei contenuti offerti da questi chatbot, ritenuti inadeguati (ad esempio perché sessualmente espliciti) per il loro grado di sviluppo cognitivo, etico, personologico.

I principi sanciti dalla disciplina privacy hanno, così, già assunto un valore determinante nella regolazione democraticamente sostenibile dei processi algoritmici, al punto da aver consentito, ad esempio alla giurisprudenza amministrativa, di rinvenirvi la disciplina di alcune determinate fattispecie e appunto, al Garante, di conformare l'utilizzo dell'i.a. con i valori propri dell'ordinamento costituzionale ed europeo e, in primo luogo, con la tutela dei diritti umani.

Di fronte all'impatto dirompente dell'I.A. sui diritti umani, si dovrebbe allora ragionare sull'opportunità di affiancare- alle Autorità di notifica e vigilanza del mercato che, ai fini dell'Artificial Intelligence Act il Governo ha indicato in Agid e Acn – un'Autorità unica che eserciti i poteri riservatele dall'art. 77, vigilando tra l'altro e con una prospettiva organica, non frammentaria, sull'incidenza dell'I.A. sui diritti umani.

Potrebbe essere questa, in particolare, **l'occasione per l'istituzione, da troppe legislature auspicata ma mai realizzata¹, dell'Autorità indipendente per la tutela dei diritti umani** con requisiti corrispondenti a quelli enunciati dai principi di Parigi (tra i quali la reale indipendenza dal Governo) e competenza estesa, tra l'altro, all'impatto sui diritti umani dell'I.A. Essa è tale da non consigliare, infatti, un'ulteriore frammentazione delle competenze tra Autorità varie e da esigere un'attribuzione unitaria per assicurare un'applicazione uniforme delle garanzie sancite sul punto dai vari plessi normativi: l'AI Act ma anche il GDPR e la direttiva 2016/680, in primis.

¹ Nonostante l'accoglimento, nella scorsa legislatura da ultimo, dell'Ordine del giorno G/2481/5/14 [già em. 13.0.1 (testo 2)] Casolati

Si potrebbe, allora, riprendere l'esame di alcuni disegni di legge già incardinati in 1^ Commissione - alcuni dei quali attribuiscono, peraltro, al Garante per la protezione dati le attribuzioni del Garante per i diritti umani - nella consapevolezza di come la rivoluzione connessa all'I.A. esiga, per la tutela dei diritti e delle libertà fondamentali sui quali incide, competenze unitarie, qualificate, indipendenti. E questo, anche in ragione dei limiti e delle aporie che la regola maggioritaria presenta, come insegnava Norberto Bobbio, di fronte a quel "territorio di frontiera" rappresentato dai diritti di libertà. Vi ringrazio.