UNIVERSITÀ
DI PAVIA

DIPARTIMENTO DI GIURISPRUDENZA
CORSO DI LAUREA MAGISTRALE IN GIURISPRUDENZA

Tutela dei diritti fondamentali e regolazione dell'Intelligenza Artificiale: profili costituzionali e comparati

Relatore: Chiar.mo Prof. Giovanni Andrea Sacco
Correlatore:

Tesi di laurea di
Chiara Serra
Matr. 510386

Anno accademico 2025/2026

Indice

Introduzione 2

Capitolo primo Principio personalistico, principio di eguaglianza e principio di trasparenza nella Costituzione italiana e in ottica comparata 4

1.1. I principi nella Costituzione italiana 4

1.2. I principi nella Western Legal Tradition 16

Capitolo secondo I principi costituzionali dinanzi agli effetti negativi prodotti dall'uso dell'IA 18

2.1. Profili di criticità dell'IA con riferimento al principio personalistico 18

2.2. Profili di criticità con riferimento al principio di eguaglianza 28

2.3. Profili di criticità con riferimento al principio di trasparenza 37

2.4. Nuove prospettive interpretative 44

Capitolo terzo Una nuova dimensione per i diritti nel settore della giustizia: rischi e benefici 50

3.1. L'impiego degli algoritmi in ambito penale: sistemi di giustizia predittiva e di polizia predittiva 50

3.2. Sistemi di *risk assessment*: COMPAS, *Public Safety Assessment* e PATTERN negli Stati Uniti 58

3.3. I rischi e i benefici dei sistemi di giustizia predittiva: *Keycrime* come esempio virtuoso in Italia 67

Capitolo quarto Un'analisi comparatistica dei modelli di disciplina 73

4.1. Il Regolamento UE 2024/1689 sull'intelligenza artificiale 73

4.2. L'*AI Action Plan* statunitense 83

Bibliografia 93

Introduzione

L'utilizzo sempre più diffuso dei sistemi di Intelligenza Artificiale porta con sé questioni inedite, cui i giuristi sono chiamati a dare risposta, spesso servendosi delle categorie del diritto tradizionale per interpretare i problemi emergenti nelle società contemporanee, a seguito della sempre maggiore pervasività delle nuove tecnologie. Si tenterà, in questa trattazione, di analizzare l'impatto dei sistemi di Intelligenza Artificiale sui diritti fondamentali, e di individuare le più recenti soluzioni legislative, dottrinali e giurisprudenziali in una prospettiva comparata.
Il primo capitolo, con un taglio introduttivo, individuerà tre principi cardine degli ordinamenti democratici, - il principio personalista, il principio di eguaglianza e il principio di trasparenza -, il cui rispetto accomuna i Paesi della tradizione giuridica occidentale.
Il secondo capitolo analizzerà l'impatto dei sistemi di *machine learning* su questi principi: fenomeni quali la profilazione, le *filter bubbles*, le *black box*, la discriminazione attraverso il *proxy*, e l'utilizzo sempre più frequente di sistemi automatizzati nei processi decisionali della Pubblica Amministrazione, possono condurre infatti ad un'erosione progressiva dell'impianto costituzionale. Si farà riferimento ai rimedi più recenti, come il *machine unlearning* e il *debiasing*, che mirano a mitigare questi rischi e a garantire la piena valorizzazione degli strumenti di Intelligenza Artificiale. Si cercherà di individuare le nuove prospettive interpretative, attraverso l'analisi delle enunciazioni di carattere programmatico ed etico più influenti a livello internazionale e, in particolare, del principio – di recente elaborazione, - dello "*human in the loop*".
Nel terzo capitolo troverà spazio un'analisi dell'utilizzo di strumenti di Intelligenza Artificiale nella giustizia penale. La diffusione di sistemi di giustizia predittiva e di polizia predittiva è dovuta agli effetti positivi che questi generano sulla prevedibilità delle decisioni e sulla prevenzione del crimine. Tuttavia l'impiego di algoritmi porta con sé notevoli rischi per il rispetto delle garanzie in ambito penale: la vigilanza sul loro funzionamento è necessaria al fine di assicurare il rispetto dei principi del giusto processo, l'obbligo di motivazione, il giudice naturale e il diritto di difesa. Si approfondiranno gli esempi più

rappresentativi, quali COMPAS, *Public Safety Assessment* e PATTERN negli Stati Uniti, con l'obiettivo di valutarli alla luce degli standard elaborati dalla giurisprudenza statunitense e attraverso la comparazione con il *software* italiano *Keycrime*.

Il quarto capitolo, di carattere più marcatamente comparatistico, analizzerà due dei principali atti di regolamentazione dell'Intelligenza Artificiale a livello globale. Si tratta del Regolamento 2024/1689 dell'Unione Europea, - l'*AI Act*, - e dell'*AI Action Plan* statunitense, emanato dall'Amministrazione Trump nel Luglio del 2025. L'analisi verterà sull'individuazione delle principali differenze tra le due soluzioni, a partire dalla diversa natura dell'atto e dal diverso grado di vincolatività, fino all'approfondimento della *ratio* dei due atti, che pare quasi diametralmente opposta. L'*AI Act*, fondato su un approccio basato sul rischio, è finalizzato alla tutela dei diritti fondamentali, mentre l'*AI Action Plan*, caratterizzato da una forte connotazione politica e da un'impostazione ingegneristica, mira a favorire lo sviluppo economico attraverso la deregolamentazione del fenomeno. La trattazione continuerà con una valutazione dei punti di forza e di debolezza dei rispettivi approcci, per poi concludersi con l'inquadramento delle prospettive auspicabili per il futuro della regolamentazione giuridica dell'Intelligenza Artificiale.

Capitolo primo
Principio personalistico, principio di eguaglianza e principio di trasparenza nella Costituzione italiana e in ottica comparata

1.1. I principi nella Costituzione italiana

Il costituzionalismo moderno si fonda sulle riflessioni politiche, storiche, filosofiche e giuridiche che hanno caratterizzato il mondo occidentale a partire dalla fine del diciassettesimo secolo. Le Costituzioni del secondo dopoguerra sono infatti l'ultimo atto di un'elaborazione dottrinale che ha radici risalenti: in particolare, sono stati momenti cruciali nella definizione del costituzionalismo le due importanti rivoluzioni di fine Settecento. La rivoluzione americana e quella francese hanno infatti condotto alla stesura, rispettivamente, della *Declaration of Independence* americana del 1776 e la *Déclaration des Droits de l'Homme et du Citoyen* francese del 1789. Il concetto di "costituzione", che emerge con forza nelle Dichiarazioni, era finalizzato alla realizzazione di una forma di governo che non fosse mera espressione dell'arbitrio politico, e all'affermazione – come condizione imprescindibile per la limitazione del potere pubblico - dei diritti della persona, a cui dovevano essere accordate idonee garanzie[1].

La Costituzione italiana, entrata in vigore il 1° Gennaio 1948, è stata approvata dall'Assemblea Costituente con il chiaro intento di condannare il passato del regime fascista e dei totalitarismi in genere, che avevano sconvolto l'Europa della prima metà del secolo scorso. "Questa frattura tra un vecchio, che si vuol seppellire, e un nuovo *ordo saeclorum*, che si vuole che nasca e viva"[2], ha permesso che la Costituzione fosse permeata di un principio che, primo fra tutti, salvaguarda dall'assorbimento nello Stato delle persone e delle strutture sociali[3]. Si tratta del principio personalista, sancito all'art. 2 della Costituzione. Esso

[1] M. Fioravanti, *Il principio di eguaglianza nella storia del costituzionalismo moderno*, in "il Mulino", a. II (1999), pp. 609-630.
[2] G. Zagrebelsky, *Trent'anni dopo. Sulle riletture del* Diritto mite, in "il Mulino", a. XLV (2025), p. 766.
[3] G. Zagrebelsky, *Trent'anni dopo. Sulle riletture del* Diritto mite, in "il Mulino", a. XLV (2025), p. 766.

recita: "*La Repubblica riconosce e garantisce i diritti inviolabili dell'uomo, sia come singolo sia nelle formazioni sociali ove si svolge la sua personalità, e richiede l'adempimento dei doveri inderogabili di solidarietà politica, economica, sociale*". Tale principio infonde di significato l'intero ordinamento costituzionale, ed è da considerarsi, seppur con le dovute differenze, espressione della comune tradizione giuridica occidentale. I diritti consacrati dall'art. 2 sono dichiarati inviolabili: emerge qui l'intenzione dei costituenti, che hanno voluto in questo modo segnare il superamento dello Stato così com'era inteso prima del referendum del 2 giugno 1946[4]. Si può affermare a ragione che il costituzionalismo contemporaneo pone come fine ultimo della Repubblica la piena realizzazione della persona, che si autodetermina con proprie scelte. A riprova della centralità del principio personalista, e del suo valore nel sistema della Costituzione, basti pensare alla previsione dell'istituto della riserva di legge rinforzata in materia di tutela dei diritti di libertà: sottraendo al principio maggioritario tali posizioni giuridiche soggettive, la riserva rinforzata pone un limite all'esercizio del potere politico, e afferma la primazia dell'uomo sull'ordinamento giuridico, sullo Stato in senso lato[5].

Il principio personalista è, come si è detto, un principio fondante degli ordinamenti giuridici occidentali, ispiratore di tutte le disposizioni a garanzia dei diritti e delle libertà fondamentali. Adottando una prospettiva interna all'articolo, si vede come sia legato a doppio filo a quello pluralistico e quello solidaristico. La partecipazione e l'integrazione, garantite dalle comunità intermedie, e l'adempimento degli inderogabili doveri di solidarietà economica, politica e sociale, sono in posizione di logica funzionalità rispetto al principio personalista. Le riflessioni sulla portata dell'art. 2 comportano però anche un'analisi delle relazioni che questo principio ha con altri principi della Costituzione, a cominciare dall'art. 1. Tale relazione emerge innanzitutto nella lettura in combinato disposto delle due norme: in primo luogo, è il lavoro il mezzo consacrato dalla Costituzione per la realizzazione dell'individuo e lo sviluppo

---

[4] P. Gori, *I diritti fondamentali*, in L. Delli Priscoli (a cura di), La Costituzione vivente, Giuffrè, Milano 2023 p. 71.

[5] A. Vedaschi, *Il principio personalista,* in L. Mezzetti (a cura di), Principî costituzionali, G. Giappichelli Editore, Torino 2011 p. 275.

della sua personalità. In questo senso il principio lavorista e il principio personalista sono da ritenersi complementari. In secondo luogo, la relazione tra i due articoli appare, forse in modo ancor più pregnante, se si guarda al carattere prodromico del principio personalista rispetto all'intero sistema democratico. Il riconoscimento e la garanzia dei diritti ex art 2 segnano la qualità stessa della democrazia, e in assenza di tale principio fondante non potrebbe aversi una Repubblica democratica. È indubbio anche il collegamento dell'art. 2 con l'art. 3 della Costituzione. È con il riconoscimento dell'uguaglianza infatti che l'essere umano può realizzarsi nella dimensione sociale, attraverso il riconoscimento della pari dignità. Essa si concreta, però, non con l'affermazione di una mera uguaglianza formale, baluardo della concezione dello Stato liberale, ma attraverso la garanzia di un'uguaglianza sostanziale, da ottenersi attraverso la rimozione da parte dello Stato degli ostacoli di ordine economico e sociale[6].
Come detto in precedenza, l'esercizio dei pubblici poteri da parte dello Stato incontra nell'art. 2 un limite, costituito dall'inviolabilità dei diritti fondamentali, il cui godimento non può essere compresso. Secondo l'orientamento prevalente, il quale adotta una lettura "forte" dell'inviolabilità, tali diritti non possono essere intaccati nel loro nucleo duro senza che con essi venga intaccata la forma di Stato. La portata dei diritti fondamentali può essere meglio compresa innanzitutto precisando che essi sono imputati *ab origine* alla persona: essi non sono creati - ma riconosciuti -, dall'organizzazione statale, e preesistono ad essa. Titolare è l'uomo, non il cittadino: i diritti inviolabili sono indipendenti rispetto al rapporto di cittadinanza, sebbene siano ammesse discipline differenziate per i non cittadini, i cui diritti di libertà "rappresentano un *minus* rispetto alla somma dei diritti di libertà riconosciuti a coloro che hanno la cittadinanza italiana"[7]. In ogni caso, nel definire la natura di questi diritti, la dottrina tradizionale ne ha qualificato i caratteri, che meglio permettono di comprenderne la rilevanza. Si tratta di posizioni giuridiche soggettive attive di carattere assoluto, inalienabili e imprescrittibili[8].

---

[6] A. Vedaschi, *Il principio personalista*, in L. Mezzetti (a cura di), Principî costituzionali, G. Giappichelli Editore, Torino 2011 pp. 275-276.
[7] Vedaschi, op. cit., p. 280.
[8] Ibid.

Secondo l'orientamento prevalente in dottrina[9], l'art. 2 sarebbe una norma di produzione giuridica. Tale linea di pensiero risponde a istanze garantistiche e di apertura nei confronti di quelli che sono sovente denominati "nuovi diritti", non positivizzati nella Carta, ai quali viene accordata copertura costituzionale attraverso la lettura della norma come clausola "aperta". Questa interpretazione non è però esente da problematiche: prima fra tutte, quale sia il fondamento di questi diritti. Una parte della dottrina ritiene che i diritti inediti siano ricavabili dalle "norme generalmente riconosciute" a livello internazionale, la cui costituzionalizzazione sarebbe legittimata dalla lettura in combinato disposto degli artt. 2 e 10 Cost. Diversa posizione, che fa leva sul verbo "*riconoscere*" ex art. 2, vede la loro fonte nel diritto naturale; altri, più saldamente ancorati al dato positivo, riconoscono la fonte di questi diritti nelle forze economiche, politiche e culturali che compongono la c.d. "costituzione materiale" in un dato momento storico. Nessuna di queste tesi appare però del tutto convincente e scevra da possibili obiezioni, e la questione del fondamento dei nuovi diritti appare ad oggi irrisolta. Una parte minoritaria della dottrina ritiene l'art. 2 una clausola "chiusa", che non ammette dunque aperture a diritti che non siano espressamente contemplati dal dettato costituzionale. Tale posizione si muove principalmente dall'intento di evitare i rischi insiti nell'orientamento opposto, criticato in quanto ricondurrebbe il fondamento di inediti diritti ad una norma, l'art. 2, che si limita a riconoscerli, non dicendo alcunché su quale sia il loro contenuto e la loro disciplina. L'esito a cui si arriverebbe, secondo l'orientamento in esame, sarebbe del tutto inutile e rischioso: da una parte, un indebolimento della rigidità della Costituzione, dall'altra un incremento del margine interpretativo – e quindi discrezionale - dei giudici, a discapito del ruolo di produzione del diritto del Parlamento[10]. Anche questo indirizzo giurisprudenziale non appare esaustivo poiché, sebbene sottolinei a ragione le criticità delle posizioni più estreme del primo orientamento di cui sopra, non tiene conto del dato sociale, della necessità che gli strumenti giuridici siano in grado di far fronte ad una realtà in continua

[9] P. Barile, *La Costituzione come norma giuridica: profilo sistematico*, Barbera, Firenze 1951, pp. 55 e ss.
[10] Vedaschi, op. cit., pp. 283-284.

evoluzione. Si pensi alla tensione, rilevante ai fini di questa dissertazione, tra i diritti e le libertà fondamentali della persona e l'impiego dell'Intelligenza Artificiale in quasi tutti gli ambiti della nostra vita, avvenuto in maniera tanto repentina quanto pervasiva. In assenza del riconoscimento alle nuove fattispecie lesive di adeguata protezione costituzionale, molte delle criticità legate a un uso incontrollato dell'IA rimarrebbero insolute. Una lettura intermedia dell'articolo in parola, e che pare convincente, distingue la tesi estensiva dall'interpretazione estensiva delle norme costituzionali. Secondo quest'orientamento, l'articolo attribuirebbe in certa misura una capacità estensiva alle norme costituzionali, senza però riconoscere un'indiscriminata apertura del catalogo dei diritti fondamentali. L'art. 2 costituirebbe quindi una legittimazione delle potenzialità espansive delle singole disposizioni costituzionali poste a tutela dei diritti tradizionali. In ogni caso, si può affermare che ad oggi la dottrina è concorde, sebbene con differenti approcci e gradazioni, nell'attribuire all'art. 2 se non una funzione propriamente creativa di diritti inediti, quantomeno una funzione estensiva delle tutele costituzionali a nuove declinazioni dei diritti tradizionalmente riconosciuti[11].

Sul piano comparatistico, si potrebbero individuare delle similitudini tra l'interpretazione dell'art. 2 come fattispecie a clausola aperta e la *living instrument doctrine*, sostenuta dalla Corte EDU a partire dalla sentenza "Tyrer c. Regno Unito" del 1978: tale indirizzo dottrinale infatti si propone di guardare al dettato costituzionale con la coscienza sociale attuale, interpretando le disposizioni in modo da garantire tutele a fattispecie non configurabili al tempo in cui il testo è stato redatto. Opposta, e sostenuta da quella parte della dottrina italiana che ritiene l'art. 2 una fattispecie a clausola "chiusa", è la lettura originalista, propria del costituzionalismo statunitense di *common-law*[12]. Tale indirizzo, come accennato, mantiene saldo l'ancoraggio al dato testuale, quale espressione della volontà dei costituenti, per cui l'evoluzione della Costituzione dovrebbe dipendere unicamente dal legislatore, nel rispetto dei confini tracciati

[11] Vedaschi, op. cit., pp. 281-285.
[12] P. Gori, *I diritti fondamentali*, in L. Delli Priscoli (a cura di), La Costituzione vivente, Giuffrè, Milano 2023 pp. 61-62.

dalle parole dei costituenti[13]. Ciò comporta evidentemente una restrizione del catalogo dei diritti inviolabili. In ogni caso, la giurisprudenza della Corte Costituzionale è ormai costante nel ritenere che il novero dei diritti fondamentali non possa essere ricondotto unicamente ai principi sanciti in Costituzione, abbracciando quindi l'interpretazione dell'art. 2 come fattispecie a clausola "aperta". Ne consegue che i diritti non espressamente contemplati in Costituzione, attinenti ai problemi più urgenti della società contemporanea, hanno comunque ottenuto protezione costituzionale, dapprima per via giurisprudenziale fino ad essere disciplinati da fonti di rango primario. Sono stati affermati in quanto declinazioni dei diritti riconosciuti dalle norme costituzionali: è il caso, ad esempio, del diritto alla privacy, del diritto all'identità personale e del diritto all'ambiente. Pur essendo l'elaborazione della Corte Costituzionale centrale nell'interpretazione dell'art. 2, un apporto significativo è da attribuire anche al giudice ordinario che, nel perimetro delle sue funzioni, ha sovente contribuito all'estensione della tutela a nuove fattispecie. Ciò è avvenuto ad esempio in materia di diritto alla reputazione personale e di diritto alla riservatezza, il cui fondamento normativo è stato ricondotto proprio all'art. 2[14].

Il principio personalista assume ad oggi un ruolo inedito, legato alla presenza di fonti esterne all'ordinamento statale che regolano numerosissime materie. Nella realtà contemporanea infatti vi è un'incessante integrazione tra ordinamenti nazionali e ordinamenti internazionali e sovranazionali: la tutela multilivello dei diritti e il dialogo tra corti che ne consegue trovano, secondo autorevole dottrina, la loro legittimazione costituzionale proprio nell'art. 2, in combinato disposto con l'art. 3 della Costituzione. Si sostiene infatti che il canone c.d. della più intensa tutela dei diritti veda il suo ancoraggio costituzionale nel principio personalista sancito dal secondo articolo del Testo costituzionale. Per questa ragione, in caso di contrasto tra disposizioni interne ed esterne all'ordinamento, sarebbe da ritenere prevalente la norma che meglio tutela la persona umana,

---

[13] C. Tripodina, *L'argomento originalista nella giurisprudenza costituzionale in materia di diritti fondamentali*, in F. Giuffrè e I. Nicotra (a cura di), Lavori preparatori ed *original intent* nella giurisprudenza della Corte costituzionale, G. Giappichelli Editore, Torino 2008 p. 248.
[14] P. Gori, op. cit., pp. 62-65.

indipendentemente da quale sia la fonte. In altre parole, se la disposizione esterna all'ordinamento nazionale accorda una maggiore protezione a diritti fondamentali, e quindi in ultima analisi alla dignità umana, questa deve prevalere rispetto alla Costituzione, e non perché imposto da un ordinamento sovranazionale, ma per previsione della Costituzione stessa, e in particolare per previsione dell'art. 2, nella parte in cui riconosce e garantisce i diritti inviolabili dell'uomo. Parte della dottrina sostiene infatti che gli stessi artt. 10 e 11 Cost. siano da ritenersi dei "veicoli" dell'art. 2, in quanto permetterebbero l'espansione delle disposizioni sancite a livello internazionale e sovranazionale nel nostro ordinamento, al fine di garantire la miglior tutela costituzionale possibile. In questo senso, l'art. 2 potrebbe costituire la "porta" del nuovo costituzionalismo contemporaneo[15].

Il principio di eguaglianza è sancito all'art. 3 Cost., il quale stabilisce al primo comma che "*Tutti i cittadini hanno pari dignità sociale e sono eguali davanti alla legge, senza distinzione di sesso, di razza, di lingua, di religione, di opinioni politiche, di condizioni personali e sociali*". Si tratta dell'enunciazione del principio di eguaglianza in senso formale, che non tiene quindi in considerazione le differenze economiche e sociali tra gli individui, ma sancisce il principio di eguaglianza davanti alla legge, il principio della pari dignità sociale e il divieto di discriminazione. Questa declinazione del principio di eguaglianza, appartenente al patrimonio del costituzionalismo moderno, è propria dello Stato liberale ed eredità delle Carte costituzionali ottocentesche. Essa segna una cesura rispetto a un ordine sociale e giuridico basato sulla divisione in ceti, prevedendo un ordinamento in cui la legge non dipende dallo status di appartenenza, ma è formulata "in termini soggettivamente universali, indifferenti alle diverse condizioni personali e sociali esistenti fra gli uomini"[16]. Molti fra i costituenti, sensibili alla necessità che il principio in parola dovesse concretarsi in un'effettiva garanzia di eguaglianza nella realtà del Paese, hanno previsto al

[15] Vedaschi, op. cit. pp., 287-288.
[16] F. Polacchini, *Il principio di eguaglianza*, in L. Mezzetti (a cura di), Principî costituzionali, G. Giappichelli Editore, Torino 2011 p. 306.

secondo comma dello stesso art. 3 il principio di eguaglianza in senso sostanziale, sebbene l'inserimento di tale disposizione fu discusso a lungo in seno all'Assemblea Costituente: la proposta suscitò opposizioni da parte dei liberali, perché visto come un superamento della concezione di Stato liberale a cui non volevano rinunciare[17]. Alla fine, prevalse l'inserimento del secondo comma nella forma che qui si enuncia: "*È compito della Repubblica rimuovere gli ostacoli di ordine economico e sociale, che, limitando di fatto la libertà e l'eguaglianza dei cittadini, impediscono il pieno sviluppo della persona umana e l'effettiva partecipazione di tutti i lavoratori all'organizzazione politica, economica e sociale del Paese*". La disposizione costituisce integrazione e sviluppo del primo comma, e ne è completamento: l'eguaglianza sostanziale infatti è prodromica all'eguaglianza formale, poiché quest'ultima non potrebbe attuarsi senza la rimozione di quegli ostacoli che di fatto impediscono l'eguaglianza fra gli uomini.

Un'interpretazione letterale del primo comma, sostenuta in passato, riteneva che la disposizione avesse come destinatari i soli cittadini. Ad oggi dottrina e giurisprudenza sono invece unanimi nell'affermare che l'enunciazione si riferisca anche agli stranieri; tuttavia occorre fare delle precisazioni in merito. La Corte Costituzionale si è espressa nel senso di riconoscere il principio di eguaglianza agli stranieri "*quando venga in gioco il riferimento al godimento dei diritti inviolabili dell'uomo*"[18], interpretando quindi la norma a sistema con l'art. 2 Cost., e chiarendo che il rispetto del principio di eguaglianza si esige per gli stranieri in riferimento alla sola tutela dei diritti fondamentali e di quei diritti che sono connessi ad un ordinamento democratico. Di conseguenza, la Corte ammette la possibilità che altri diritti vengano riconosciuti solo ai cittadini e non anche agli stranieri. In questo senso, la giurisprudenza costituzionale è omogenea nell'interpretazione con riferimento agli artt. 2 e 3. Destinatari della disposizione ex art. 3 sono anche le persone giuridiche.

---

[17] C. Bernardo, *Il principio di uguaglianza*, in L. Delli Priscoli (a cura di), La Costituzione vivente, Giuffrè, Milano 2023 p. 103.
[18] Polacchini, op. cit., p. 311.

Delle quattro enunciazioni contenute nell'articolo, tre sono al primo comma. La statuizione che occorre analizzare per prima è contenuta nell'espressione "*tutti i cittadini ... sono eguali davanti alla legge*". L'eguaglianza davanti alla legge è un principio fondante dell'intero ordinamento, innanzitutto perché da esso dipende l'esistenza stessa di uno Stato democratico. Il legislatore è, sulla base di tale principio, vincolato a imporre norme generali e astratte. La posizione assunta da Carlo Esposito, uno dei primi interpreti dell'art. 3, fa proprio riferimento alla generalità e all'astrattezza delle norme per la realizzazione della parità giuridica. Egli infatti sosteneva che la disposizione "*definisce in modo incontrovertibile la forza e l'efficacia*"[19] della legge, non ne disciplina il contenuto. Il principio di eguaglianza consisteva e assumeva rilevanza concreta, secondo questo indirizzo dottrinale, nel semplice divieto di leggi personali, che assicuravano la parità giuridica fra gli individui. Nell'attuale forma di Stato, pluralista e quindi portatrice di istanze di differenziazione, l'eguaglianza intesa come pura generalità e astrattezza della legge porta però con sé il rischio di discriminazioni. Non è possibile infatti immaginare delle leggi uguali per tutti senza che sia attuato l'opportuno discrimine tra situazioni che presentano elementi di differenziazione, e che quindi richiedono un differente trattamento. Per questo, l'art. 3 è da interpretarsi alla luce di un'istanza di giustizia, da affiancare all'istanza di universalizzabilità: nell'ottica di realizzazione dell'eguaglianza sostanziale, occorre servirsi non solo del vaglio di costituzionalità, ma anche di un controllo di ragionevolezza, volto ad assicurare la razionalità dell'ordinamento nel suo complesso[20].

L'art. 3 contiene nello stesso enunciato, oltre all'affermazione del principio di eguaglianza davanti alla legge, anche quello della pari dignità sociale. Il riferimento non è da ricondurre a un generico richiamo alla dignità umana, né è da intendersi come ripetizione enfatica del concetto di eguaglianza davanti alla legge, sebbene la stessa Corte Costituzionale abbia utilizzato a più riprese le due espressioni come interscambiabili, e quindi equivalenti. Tuttavia è accaduto che la Corte meglio esplicitasse come sia da intendersi la nozione di pari dignità

[19] Polacchini, op. cit., p. 308.
[20] Polacchini, op. cit., p. 309

sociale: essa "*esprime un principio generale che da un lato importa l'illegittimità di tutte le misure legislative che collegano particolari distinzioni di rilevanza sociale a circostanze che non siano dipendenti da capacità e meriti personali, dall'altra concorre ad interpretare le stesse norme costituzionali nel senso più rispettoso di siffatta esigenza*"[21].

Il fondamento costituzionale del principio di trasparenza è da ricondurre all'articolo 97 della Costituzione, in particolare al secondo comma. Esso recita: "*i pubblici uffici sono organizzati secondo disposizioni di legge, in modo che siano assicurati il buon andamento e l'imparzialità dell'amministrazione*". L'evoluzione dottrinale e giurisprudenziale ha portato a individuare nella trasparenza un elemento prodromico al buon andamento e all'imparzialità.
Il combinato disposto degli artt. 3 e 97 della Costituzione individua nell'imparzialità un principio guida della Pubblica Amministrazione nell'esercizio delle sue funzioni: questa, però, può essere garantita unicamente a condizione che sia assicurata la trasparenza nel procedimento. In questo senso, la trasparenza assume un ruolo centrale quale presidio dell'imparzialità: essa è da intendersi come conoscibilità e accessibilità delle motivazioni dell'atto, senza che il processo decisionale dell'Amministrazione sia caratterizzato da opacità che rendano queste ultime imponderabili.
Il buon andamento è garantito dal controllo sociale sull'attività dei funzionari pubblici, che conferisce legittimità all'attività amministrativa e democraticità all'intero sistema. Tale principio è volto ad assicurare l'efficienza, l'economicità e l'efficacia dell'Amministrazione, e pone in capo all'autorità pubblica l'onere della responsabilità nell'esercizio delle sue funzioni, al fine di scoraggiare la cattiva gestione e la corruzione. Affinché questo si realizzi occorre che venga rispettato il principio di trasparenza, condizione necessaria affinché sia garantito un controllo diffuso sull'operato della Pubblica Amministrazione.
Il principio di trasparenza ha ottenuto un riconoscimento positivo molto più tardi, nella legge n. 241/1990 sul procedimento amministrativo. Quest'ultima,

[21] Ibid.

nel ridefinire il rapporto fra amministrazione e cittadini, - a favore di questi ultimi, non più meri destinatari delle decisioni, - individua, all'art. 1 co. 1, la trasparenza quale principio generale dell'attività amministrativa. Il legislatore del Novanta dedica il Capo V all'accesso ai documenti amministrativi, ed è in questa sede che principio di trasparenza trova concreta declinazione in una serie di disposizioni: la loro rilevanza è tale che la l. n. 241/1990 è anche conosciuta come legge sull'accesso agli atti o sulla trasparenza[22]. L'art. 22, comma 2[23], è una norma cardine in questo senso. Essa dispone che "l'accesso ai documenti amministrativi, attese le sue rilevanti finalità di pubblico interesse, costituisce principio generale dell'attività amministrativa al fine di favorire la partecipazione e di assicurarne l'imparzialità e la trasparenza".

Si tratta ancora, però, di una trasparenza limitata, il cui limite è sancito già al primo comma, che precede persino il principio generale: il diritto di accesso è riconosciuto unicamente agli "interessati", il cui interesse sia diretto, concreto e attuale, e che si riferisca ad una situazione giuridicamente tutelata e collegata al documento al quale è chiesto l'accesso[24]. Ancora, le istanze d'accesso non sono ammesse se volte ad un controllo generalizzato delle pubbliche amministrazioni[25].

Un primo cambio di rotta si ha con l'affermazione del principio di accessibilità totale, ad opera del d.lgs. n. 150 del 2009. L'articolo 11 modifica la nozione di trasparenza, ora intesa come al servizio non solo del destinatario dell'atto, ma della collettività, al fine di garantire quelle "forme diffuse di controllo del rispetto dei principi di buon andamento e di imparzialità"[26], cui si è fatto cenno. Il presupposto è quello di un'amministrazione "aperta", orientata all'*open data*: è con il legislatore del 2009 che il principio di trasparenza assume un significato più preciso. In particolare, così come intesa nel d.lgs. n. 150, la trasparenza può dirsi finalizzata all'efficienza e al miglioramento dei servizi pubblici, e in questo senso opera sul versante della *performance* esterna dell'amministrazione; d'altra

---

[22] S. Talamo, *La trasparenza totale. Partenza lenta ma grandi orizzonti*, Smart24 PA+, Gruppo24Ore, Milano 2019, p. 4.
[23] Così modificata dall'articolo 10, comma 1, legge n. 69/2009.
[24] Talamo, op. cit., p. 4.
[25] Legge n. 241/1990, art. 24 co. 3.
[26] D.lgs. n. 150 del 2009, art. 11 co. 1.

parte, opera anche nel senso di prevenire la corruzione e la cattiva amministrazione in genere, e di conseguenza può dirsi finalizzata alla responsabilizzazione dell'autorità pubblica[27].

La trasparenza viene intesa più specificatamente come "audit-anticorruzione"[28] con il d.lgs. n. 33 del 14 Marzo 2013. Oltre a riordinare in un *corpus* unitario, sistematico e semplificato le disposizioni disseminate nell'ordinamento, il decreto legislativo prevede ulteriori obblighi di informazioni specifici, atti a contrastare il fenomeno della *maladministration* in genere. La disposizione di assoluta novità riguarda però il riconoscimento di un diritto di accesso civico alle informazioni e ai dati in relazione ai quali non è stato adempiuto dall'amministrazione l'obbligo di pubblicità. Tale diritto soggettivo costituisce un implemento rispetto al generale diritto di accesso, attivabile da chiunque[29]. Oggetto di regolazione del decreto legislativo è stato anche il problematico rapporto tra la tutela dei dati personali e il principio di trasparenza: il legislatore ha tentato di operare un bilanciamento in questo senso, attraverso l'applicazione del principio di proporzionalità[30].

Il principio di trasparenza amministrativa ha visto un ulteriore fase, ancor più avanzata, con il d.lgs. del 25 Maggio 2016, n. 97: esso si ispira ai principi del *Freedom of Information Act* statunitense[31], in cui la regola è "la *full disclosure* di ogni atto pubblico"[32]. Il *F.O.I.A.* italiano è l'esito di un'evoluzione iniziata a partire dalla legge del Novanta e che ha visto un progressivo potenziamento del principio di trasparenza, fino al raggiungimento, con la trasparenza digitale, all'accesso universale alle informazioni relative all'attività dell'amministrazione pubblica.

---

[27] F. Patroni Griffi, *La trasparenza della pubblica amministrazione tra accessibilità totale e riservatezza*, in federalismi.it n. 8/2013, Roma 2013, p. 4.
[28] Talamo, op. cit., p. 5.
[29] B. Ponti, *Il codice della trasparenza amministrativa: non solo riordino, ma ridefinizione complessiva del regime della trasparenza amministrativa online,* in www.neldiritto.it, 2013.
[30] Patroni Griffi, op. cit., p. 10.
[31] Si tratta di una legge sulla libertà di informazione su atti delle amministrazioni pubbliche, emanata negli Stati Uniti il 4 Luglio del 1966 dal Presidente Lyndon Johnson.
[32] P. Falletta, *Il* freedom of information act *italiano e i rischi della trasparenza digitale*, in federalismi.it, n. 23/2016, p. 2.

Ad oggi, l'impiego di sistemi di Intelligenza Artificiale nella Pubblica Amministrazione rischia di minacciare il rispetto del principio di trasparenza, per via dell'opacità che spesso caratterizza i processi decisionali degli algoritmi di *machine learning*, e la conseguente impossibilità per il cittadino di conoscere le ragioni alla base dell'atto di cui è destinatario.

1.2. I principi nella Western Legal Tradition

Il costituzionalismo contemporaneo ha le sue radici nella tradizione politica, filosofica e giuridica occidentale. Principi quali lo stato di diritto, la separazione dei poteri, il principio di legalità, la libertà individuale, l'eguaglianza, il principio rappresentativo, la tutela giurisdizionale dei diritti e l'indipendenza dei giudici sono frutto di un'elaborazione dottrinale che ha visto la sua espressione nella nascita, tra la fine del Settecento e la prima metà dell'Ottocento, dello Stato liberale[33].

Con l'espressione Stato liberale si intende quella forma di Stato caratterizzata da un'ideologia liberista e individualista[34], fondata sulla nozione di Stato minimo, che pare distante dai caratteri degli Stati occidentali contemporanei, ma che costituisce comunque un'eredità di principi che ancora oggi fondano e caratterizzano gli Stati della Western Legal Tradition. È con lo Stato liberale, infatti, che nasce la nozione di Stato di diritto, principio cardine di tutte le moderne democrazie occidentali. Le sue radici risiedono nelle teorie filosofiche del giusnaturalismo e del contrattualismo, che hanno condotto alla sostituzione del potere sovrano che si legittima da sé con un potere politico limitato. Sono i cittadini, che nascono uguali, a decidere liberamente di delegare il potere ad un'autorità: da qui discende il fondamento della legittimità dello Stato, che risiede nel contratto sociale.

Lo Stato assume, quale finalità politico costituzionale, quella garantista: in particolare, è lo Stato a presidiare la tutela delle libertà e dei diritti degli individui[35]. Ciò avviene attraverso la legge: ogni limitazione della libertà dell'individuo non può avvenire se non tramite disposizione di legge. Essa deve

---

[33] R. Bin, G. Pitruzzella, *Diritto Costituzionale*, G. Giappichelli Editore, Torino 2020, p. 43 e ss.
[34] Ibid.
[35] Bin, Pitruzzella, op. cit., p. 43.

essere generale e astratta, ed espressione dai rappresentanti della Nazione, che vengono eletti dal corpo elettorale ed operano in Parlamento[36].

Si afferma dunque, - ad imitazione della forma di governo inglese, - il modello parlamentare, che è senza dubbio il più diffuso nell'Occidente, insieme al modello presidenziale[37].

Al di là delle pur molto significative differenze, la comune storia occidentale ci consegna un'eredità, una tradizione giuridica che ancora oggi è alla base delle moderne democrazie occidentali. Questo substrato condiviso permette di muovere una riflessione di ampio respiro sulla tenuta di alcuni di questi principi fondamentali di fronte alle nuove sfide dell'epoca digitale, e un'analisi comparatistica di alcune delle soluzioni regolatorie proposte a livello mondiale.

[36] A. Buratti, *Western Costitutionalism. History, Institutions, Comparative Law,* Springer e G. Giappichelli Editore 2023, p. 9 (versione online).
[37] P. G. Lucifredi, *Almagesto costituzionale*, Giuffrè Editore, Milano 1998, p. 137.

Capitolo secondo

I principi costituzionali dinanzi agli effetti negativi prodotti dall'uso dell'IA

2.1. Profili di criticità dell'IA con riferimento al principio personalistico

L'apertura al web per chiunque possedesse un computer, e la nascita di alcune aziende che si sarebbero trasformate in multinazionali come Google, hanno portato con sé la nascita di una nuova forma di capitalismo, i cui imperativi economici erano già delineabili a partire dalla fine degli anni Ottanta del secolo scorso. Shoshanna Zuboff, nel suo libro "*Il capitalismo della sorveglianza*", individua le esigenze che con vigore si sono imposte sin dagli albori dell'epoca digitale, facendo riferimento all'"estrazione" e alla "predizione" dei dati, e all'elaborazione di "metodi per la modifica del comportamento"[38].

Gli ingegneri di Google capirono ben presto che dalle ricerche effettuate dagli utenti era possibile ricavare una serie di dati, che potevano essere reinvestiti per migliorare la funzione di ricerca. Questa prima fase viene definita "ciclo di reinvestimento del valore comportamentale"[39]: Google utilizzava i dati raccolti tramite il suo motore di ricerca per rendere i risultati sempre più attendibili e precisi. Si tratta insomma dello stesso meccanismo che regola l'IA: l'incremento di efficienza è determinato dall'esposizione a una quantità sempre maggiore di dati, da cui l'Intelligenza Artificiale apprende. Fino al 1999 dunque, il valore comportamentale - cioè l'insieme dei dati forniti dagli utenti attraverso le ricerche che essi effettuavano, - veniva utilizzato a esclusivo vantaggio di questi ultimi, in quanto grazie ad esso ottenevano un servizio sempre migliore.

A partire dal Duemila però, per far fronte all'esigenza di trasformare gli investimenti in ricavi, Google iniziò a servirsi di quei dati per altri usi. Ciò emerge con chiarezza da un brevetto registrato nel 2003: "*Generating User Information for Use in Targeted Advertising*"[40]. Il valore comportamentale viene usato per far combaciare l'*advertising* agli interessi di un determinato utente, deducibili dai dati relativi al suo comportamento online. La quantità di dati

---

[38] S. Zuboff, *Il capitalismo della sorveglianza: il futuro dell'umanità nell'era dei nuovi poteri*, Luiss University Press, Roma 2019 p. 77.
[39] Zuboff, op. cit., p. 79.
[40] Zuboff, op. cit., p. 87.

raccolti non è più quella sufficiente a migliorare il servizio: è molto maggiore. Questa seconda fase, in cui la scoperta del surplus comportamentale segna la nascita di un sistema governato dalla logica dell'accumulazione dei dati, è cruciale per comprendere l'impatto che ha oggi l'Intelligenza Artificiale sui principi portanti di ogni società democratica, primo fra tutti il principio personalistico. La profilazione, resa possibile dall'enorme quantità di dati che gli utenti forniscono al capitalismo della sorveglianza quale materia prima, permette di predire i loro comportamenti futuri[41].

Interessa, in riferimento al contenuto del brevetto, la presenza di nuovi set di dati, gli *user profile information* ("UPI"), la cui compilazione non costituiva un ostacolo al diritto degli utenti a decidere quali informazioni personali fornire. Nel brevetto sono elencate in via esemplificativa le "tracce" lasciate dall'utente attraverso le sue ricerche che permettono di accumulare sempre più surplus comportamentale, anche in assenza di un esplicito consenso da parte dello stesso; appare verosimile ritenere che, ad oggi, le modalità di aggiramento del rifiuto dell'utente di condividere i suoi dati siano ancor più sofisticate[42].

Come si vede, il modello inaugurato da Google porta con sé l'elevato rischio che venga valicato, ogni qualvolta sia possibile, il limite costituito dal diritto all'autodeterminazione di ogni individuo. La deduzione e l'estrazione dei dati comportamentali anche in assenza del consenso dell'utente costituisce di fatto una violazione della privacy, "nascosta": i dirigenti di Google argomentavano infatti che, non trattandosi di vendita dei dati stessi, ma di vendita dei prodotti predittivi elaborati a partire dai dati, la privacy degli utenti non era stata violata[43].

Nasce quindi un nuovo mercato di vendita dei comportamenti futuri: le aziende capitaliste del web estraggono la materia prima dagli utenti, - il surplus comportamentale -, che incrementa l'intelligenza della macchina, e a sua volta la macchina produce prodotti predittivi sempre più completi e accurati, che vengono venduti agli inserzionisti, i quali mirano ad aumentare i tassi di *click through* per ogni inserzione indirizzata ad un determinato utente[44]. Tale sistema,

---

[41] Zuboff, op. cit., p. 104.
[42] Zuboff, op. cit., p. 89.
[43] Zuboff, op. cit., p. 106.
[44] Zuboff, op. cit., p. 106.

se non adeguatamente illuminato da principi che fungano da argine alle mire capitalistiche, ha un impatto profondo e radicale sulla società, le cui necessità, convinzioni, ambizioni e desideri possono essere abilmente manipolate dalle *Big Tech*.

Tra le numerosissime dichiarazioni che hanno cercato di far luce su questa e altre questioni, le più autorevoli per attualità, rilevanza, reputazione e influenza sono sei[45]. Esse sono accomunate dall'adozione degli stessi principi etici propri della bioetica, ma le cui categorie possono essere facilmente riferite anche all'ecosistema digitale. Si tratta delle categorie della beneficenza, della non maleficenza, dell'autonomia e della giustizia[46]. Alla luce di queste, si auspica un impiego dell'IA che miri al benessere dell'umanità e del pianeta, che scansi i rischi di un suo uso malevolo, - per esempio in violazione della privacy personale -, che non si appropri del diritto degli individui di "decidere di decidere", e che promuova la solidarietà e l'eguaglianza tra gli individui[47].

Pur non avendo rilevanza giuridica in senso stretto, queste enunciazioni programmatiche possono aiutare a comprendere come dovrebbe orientarsi la giurisprudenza nel misurare la tenuta dei principi costituzionali dinanzi alle nuove questioni poste dall'Intelligenza Artificiale nella società contemporanea. Prodromica ai caratteri della beneficenza, non maleficenza, autonomia e giustizia dell'IA, e quindi elemento centrale nel misurare la bontà dell'impiego che viene fatto di questa tecnologia, vi è l'esplicabilità. La valutazione del rispetto dei quattro principi etici di cui sopra non può infatti prescindere da una condizione primaria, cioè la decifrabilità dell'operato della macchina. Il processo decisionale adottato dall'IA deve essere ripercorribile, il percorso

---

[45] Per una rassegna delle serie di principi etici più influenti per l'IA si rimanda a L. Floridi, *Etica dell'intelligenza artificiale: sviluppi, opportunità, sfide*, Raffaello Cortina Editore, Milano 2022, pp. 94-95. Si tratta dei *Principi di Asilomar per l'IA* (*Future of Life Institute*, 2017), della *Dichiarazione di Montréal per l'IA responsabile* (Università di Montréal, 2017), dei principi elaborati in occasione della seconda versione di *Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Autonomous and Intelligent Systems* (IEEE, 2017), e quelli offerti nella *Dichiarazione su intelligenza artificiale, robotica e sistemi autonomi* (EGE, 2018), nonché i "cinque principi generali per un codice di intelligenza artificiale" contenuti nel rapporto del Comitato per l'intelligenza artificiale della Camera dei Lord del Regno Unito, denominato "*AI in the UK: ready, willing and able?*", e infine dei *Principi di partenariato sull'IA* (*Partnership on AI*, 2018).

[46] Ibid.

[47] Floridi, op. cit., pp. 96-100.

logico manifesto: l'algoritmo, insomma, deve essere conosciuto[48]. Questo è possibile fino a che la verifica *ex post* dell'uomo sull'operato della macchina non si renda irrealizzabile per via dello spettro di opacità di una *black box*[49], di cui si tratterà in seguito.

Come accennato, una delle questioni più dibattute e problematiche legate al *machine learning* è il tema della profilazione. Al legislatore è chiesto di regolare tale attività affinché avvenga nel rispetto del principio personalistico, e siano tutelati quindi il diritto all'autodeterminazione, la privacy dei dati nella rete, la libertà di espressione e di informazione. La profilazione è la raccolta di un enorme numero di informazioni personali, che vengono elaborate ed incrociate per fornire prodotti e servizi sempre più personalizzati, fino ad anticipare e orientare comportamenti e decisioni dell'individuo[50]. Ciò ha un impatto molto positivo sul piano economico: l'alta probabilità di riuscire a influenzare il comportamento di una certa persona, attraverso l'invio di un determinato messaggio in un determinato momento, è stato visto - già ai tempi della nascita di Google, - come il Sacro Graal della pubblicità[51]. Tuttavia la profilazione, in ragione della sua sempre maggiore accuratezza e del suo carattere pervasivo, costituisce anche un rischio per il diritto di ogni utente all'autodeterminazione, intesa come diritto dell'individuo a definire in modo autonomo e consapevole le proprie scelte esistenziali[52].

Il paradosso è che nello spazio sconfinato del web, che dovrebbe offrire una potenzialmente illimitata libertà di scelta e di informazione, quest'ultima sia vanificata da una altrettanto grande difficoltà di esercitarla[53]. L'idea tanto radicata in passato, per cui Internet potesse mettere in contatto il cittadino con la più ampia gamma di idee e opinioni, restaurando quel pluralismo

---

[48] Floridi, op. cit., p. 100.
[49] D. Messina, *La tutela della dignità nell'era digitale: prospettive e insidie tra protezione dei dati, diritto all'oblio e Intelligenza Artificiale*, Editoriale Scientifica, Napoli 2023 pp. 142-143.
[50] D. Messina, op. cit., p. 134.
[51] S. Zuboff, op. cit., p. 88.
[52] D. Messina, op. cit., p. 134.
[53] M. Manetti, *Pluralismo dell'informazione e libertà di scelta*, in "AIC – Associazione Italiana dei Costituzionalisti", 1 (2012), p. 6.

dell'informazione che si riteneva ormai non più garantito dalla televisione[54], è ad oggi da ritenere quantomeno ridimensionabile.

Un primo profilo del diritto all'autodeterminazione informativa è quello "attivo": esso può essere definito come il diritto ad essere informati, la cui tenuta è messa a rischio dalle nuove tecnologie[55]. Tra i fenomeni più significativi in questo senso, vi è quello della c.d. *filter bubble*: si tratta di uno degli esiti dell'attività di profilazione operata dai sistemi di IA. La *filter bubble* mina la capacità dell'individuo di elaborare scelte proprie, formulare propri convincimenti, e, in ultima analisi, delineare la propria personalità, poiché tali caratteri fondamentali dello sviluppo esistenziale sono in vario modo indotti e influenzati dall'algoritmo. La profilazione sempre più dettagliata, grazie alla capacità della macchina di individuare relazioni tra le migliaia di informazioni che l'utente lascia in rete[56], porta alla progressiva "chiusura su sé stessi"[57] degli utenti, che sul web entrano in contatto in misura sempre minore, se non quasi nulla, con ciò che non sia in linea con le proprie preferenze. Questo meccanismo reca enormi vantaggi economici alle attività commerciali, perché permette che un determinato prodotto o servizio arrivi a quello specifico utente che potrebbe acquistarlo, ma ha un costo altrettanto elevato.

La contropartita riguarda innanzitutto il conflitto di tali sistemi di profilazione con il principio personalistico, richiamato all'inizio di questa trattazione, per la violazione del diritto di ogni individuo al libero sviluppo della propria personalità. Tuttavia, non si può ignorare l'impatto che la profilazione ha non solo sul singolo e sulla sua capacità di autodeterminarsi, ma anche sulle società nel loro complesso, innanzitutto sulle democrazie. Il pluralismo delle idee, principio cardine di ogni società democratica, deve fare i conti con la presenza

---

[54] Manetti, op. cit., p. 6.

[55] A. Papa, *La problematica tutela del diritto all'autodeterminazione informativa nella* big data society, in *Liber amicorum* per Pasquale Costanzo – Diritto costituzionale in trasformazione, vol. I – Costituzionalismo, Reti e Intelligenza Artificiale, Collana di studi di *Consulta OnLine*, 3, 2020, Genova, p. 478.

[56] D. Messina, op. cit., p. 259.

[57] Ibid.

di "bolle epistemiche"[58] online, in cui i contenuti proposti ad ogni utente ne rispecchiano le preferenze, i bisogni, e le ideologie[59].

Alla medesima logica di "selezione per omissione delle informazioni, conoscenze, relazioni e valori"[60] risponde un altro fenomeno, quello delle *echo-chambers*. Nel caso delle *echo-chambers*, la distanza con sistemi valoriali che l'utente non condivide è, a differenza delle *filter bubbles*, volontaria e ricercata dal soggetto. È qui che il fenomeno delle *fake news* trova un terreno fertile e una cassa di risonanza, determinando una sempre più feroce radicalizzazione delle idee, che indebolisce la dimensione di realizzazione individuale, nonché collettiva, dell'essere umano.

Il secondo profilo del diritto all'autodeterminazione informativa, quello "passivo", consiste nel diritto a determinare la propria proiezione sociale[61]. Visti i rischi per la privacy determinati dall'impiego delle nuove tecnologie, si sono rese urgenti, nell'ambito dell'Unione Europea, le regolamentazioni in materia di protezione dei dati e di Intelligenza Artificiale.

Si tratta rispettivamente del GDPR (Regolamento UE 2016/679) e dell'*AI Act* (Regolamento UE 2024/1689). Mentre dell'*AI Act* si tratterà in seguito, è possibile sin da ora soffermarsi su alcune considerazioni relative al *General Data Protection Regulation* (GDPR). Per comprendere la ratio del regolamento, appare opportuno guardare alla direttiva che l'ha preceduto, la *Data Protection Directive* (DPD). La DPD è stata adottata nel 1995, ed era orientata alla responsabilità dei singoli Stati membri in materia di protezione dei dati: era demandata a questi la scelta delle misure e delle sanzioni da adottare. Ciò portò a una sostanziale ineffettività della direttiva, sebbene i meccanismi di tutela in essa contenuti fossero più stringenti di quelli previsti dall'attuale regolamento[62]. L'esito insoddisfacente portò con sé come naturale conseguenza la scelta del legislatore europeo, vent'anni dopo, di destinare la regolazione della materia a un diverso tipo di atto, cioè il regolamento 2016/679.

---

[58] D. Messina, op. cit., p. 259.
[59] D. Messina, op. cit., p. 264.
[60] D. Messina, op. cit., p. 260.
[61] Papa, op. cit., p. 480.
[62] F. Dal Monte, *The philosophy of the GDPR*, in E. Longo, A. Pin, F. Viglione (a cura di), Data protection in context: between privacy and AI, Giuffrè, Milano 2025, pp. 50-51.

Esso aveva come obiettivo quello di armonizzare la disciplina degli Stati in materia di dati personali, in netta controtendenza rispetto all'atto precedente. In quegli anni infatti era emersa con chiarezza la non trascurabilità degli effetti che le nuove tecnologie avevano sulla società, al punto che il diritto alla protezione dei dati fu espressamente previsto dalla Carta dei diritti fondamentali dell'Unione Europea, all'art. 8.

Il riconoscimento di questo diritto nella Carta di Nizza incoraggiò il passaggio dal DPD, in cui il livello di armonizzazione della materia tra gli Stati membri era basso, al GDPR, che mirava invece a rendere omogenea la disciplina in tutti gli Stati membri. Nel GDPR viene adottato un regime di "*ex ante compliance*". Si tratta di un approccio preventivo, che subordina l'ingresso delle aziende nel mercato al rispetto da parte delle stesse degli standard previsti dall'Unione in materia di protezione dei dati personali[63]. Ciò che emerge, quale segno di cambiamento rispetto alla disciplina precedente, è che il GDPR adotta una nozione di privacy strettamente legata all'individuo e al suo diritto all'autodeterminazione. Viene così superata la concezione negativa della privacy, che mira alla tutela della libertà più che della dignità, e che si concretizza nella richiesta di strumenti che prevengano le intrusioni nella sfera privata dell'individuo[64]. Ciò avviene in favore dell'opposta concezione positiva, in cui il diritto alla protezione dei dati è sancito in quanto declinazione del rispetto della dignità umana[65].

Nel regolamento del 2016, il concetto di profilazione è definito all'art. 4 come una forma di trattamento automatizzato dei dati personali, che ha come fine la valutazione di determinate caratteristiche personali relative a una persona fisica, "per analizzare o prevedere alcuni aspetti riguardanti la sua vita privata"[66].

L'avvento dell'Intelligenza Artificiale ha portato infatti con sé una rinnovata attenzione del legislatore, sia nazionale che europeo, verso il tema della protezione dei dati: l'applicazione di principi tradizionali in materia, come

---

[63] Dal Monte, op. cit., pp. 50-51.
[64] Ibid.
[65] *Carta dei diritti fondamentali dell'Unione Europea* (2012/C 326/02), in Gazzetta ufficiale dell'Unione europea, art. 1.
[66] I. M. Alagna, op. cit., p. 279.

l'*accountability* e la limitazione della finalità di trattamento, è infatti resa più complessa dall'utilizzo di questa tecnologia[67].

L'*accountability*[68] consiste nella responsabilizzazione degli operatori, i quali devono dimostrare la concreta adozione di misure finalizzate ad assicurare l'applicazione del regolamento. I titolari possono decidere in autonomia modalità, garanzie e limiti del trattamento dei dati personali, ma a condizione che questi siano stabiliti nel rispetto delle disposizioni normative e di alcuni criteri specifici indicati nel regolamento. La limitazione della finalità di trattamento[69] consiste invece nell'obbligo di assicurare che eventuali trattamenti successivi dei dati non siano incompatibili con la finalità della raccolta dei dati stessi[70]. Questi non possono, insomma, essere utilizzati per finalità diverse e ulteriori rispetto a quelle dichiarate.

Tuttavia gli standard in materia di privacy sanciti dal legislatore europeo, di cui si è fatto un breve accenno, entrano in collisione con il ruolo sempre crescente dell'Intelligenza Artificiale, che spesso ne rende opaca e complessa l'applicazione[71]. È ormai chiaro per i *policymakers* e i legislatori che la disciplina sulla protezione dei dati non può essere ritenuta sufficiente se non integrata e adattata ai nuovi sistemi di *machine learning*, e ciò innanzitutto perché l'IA è uno degli ambiti in cui la tutela della privacy è più a rischio. In particolare, l'ultimo sviluppo dell'Intelligenza Artificiale, cioè l'IA generativa, ha un effetto dirompente sulla protezione dei dati personali. Si pensi al fatto che i dati estratti da pagine web, social media, blog, forum e recensioni di prodotti e servizi vengono utilizzati per "addestrare" l'IA generativa.

Si tratta di dati personali di vario tipo condivisi dagli utenti sul web: possono essere immagini, video, testi, informazioni di contatto. Lo sviluppo e l'incremento di efficienza dell'IA, più precisamente, dipende dai dati che gli utenti lasciano in rete, e ciò porta con sé la necessità che vengano predisposti i

---

[67] E. Longo, *From data protection to AI regulation*, in E. Longo, A. Pin, F. Viglione (a cura di), Data protection in context: between privacy and AI, Giuffrè, Milano 2025, p. 13.

[68] *Regolamento (UE) 2016/679*, in Gazzetta ufficiale dell'Unione europea, artt. 5, 23-25, in particolare, e l'intero Capo IV del Regolamento.

[69] *Regolamento (UE) 2016/679*, in Gazzetta Ufficiale dell'Unione Europea, art. 5, par. 1, lett. b.

[70] *Principi fondamentali del trattamento*, in www.garanteprivacy.it (consultato il 25 Maggio 2026).

[71] Longo, op. cit., p. 13.

principi legali a cui l'utilizzo di tale strumento deve attenersi, alla luce del valore primario della dignità umana[72].

Con riferimento al tema del rispetto della dignità umana e del trattamento dei dati personali, rileva ai fini di questa trattazione il rapporto tra i sistemi di *machine learning* e il diritto all'oblio. L'espressione "*droit à l'oubli*" è stata coniata in Francia, per indicare il diritto a che le vicende di dominio pubblico legittimamente rese note non venissero nuovamente diffuse, dopo che fosse trascorso un periodo di tempo considerevole. Dall'iniziale elaborazione dottrinale francese ad oggi, lo scenario e il significato di tale diritto è cambiato notevolmente, in ragione dell'evoluzione tecnologica.

Ad oggi infatti ciò a cui si fa riferimento è non già il lasso di tempo trascorso tra il fatto e la ripubblicazione della vicenda, ma il tempo di permanenza dell'informazione nel web[73]. Internet "trattiene" i dati potenzialmente per sempre; è, possiamo dire, il luogo a cui l'umanità sta consegnando la sua eredità. Il problema della gestione dei contenuti in rete è dunque fondamentale, "sia per il diritto di informare e di essere informati, sia per l'elaborazione di quel patrimonio culturale comune che contribuisce a formare le identità personali e collettive, sia, da ultimo, per la stessa democrazia"[74].

A riprova dell'importanza assunta dal diritto all'oblio, vi è il suo autonomo riconoscimento all'interno del GDPR, all'art. 17[75]. Ancor prima del regolamento, però, era stata la Corte di Giustizia a sancire la responsabilità dei motori di ricerca con riferimento alla permanenza delle informazioni dei soggetti in rete, nella celebre sentenza *Mario Costeja Gonzàlez vs Google Spain*[76] del 13 Maggio del 2014. In quell'occasione la Corte ha sostenuto che l'attività svolta dai motori di ricerca attraverso l'indicizzazione, - cioè l'estrazione, la registrazione e l'organizzazione dei dati raccolti – rientra nel concetto di

---

[72] Longo, op. cit. pp. 15-16.

[73] I. M. Alagna, *Diritto all'oblio e mancato oscuramento dei dati sensibili: responsabilità, profili sanzionatori e rimedi pratici applicabili*, Giuffrè, Milano 2023, pp. 4-5.

[74] Alagna, op. cit., p. 3.

[75] *Regolamento (UE) 2016/679*, in Gazzetta Ufficiale dell'Unione Europea. All'art. 17 è sancito il diritto dell'interessato di ottenere dal titolare del trattamento la cancellazione dei dati personali che lo riguardano senza ingiustificato ritardo.

[76] Si tratta della sentenza pronunciata dalla Corte di Giustizia Europea, nella causa C-131/12, conosciuta come *Google Spain* (2/2014).

trattamento, anche se le informazioni raccolte non sono qualificate come dati personali. La pubblicazione di informazioni, anche se non personali e anche se già pubblicate in precedenza dai media, non è quindi da ritenere esente da responsabilità da parte dei motori di ricerca. Nella pronuncia, la Corte ha anche precisato che se le informazioni su un soggetto possono essere reperite inserendo il suo nome nel motore di ricerca, il gestore è tenuto, in presenza di determinate condizioni, ad eliminare i *link* a pagine web pubblicate da terzi contenenti informazioni relative a tale persona[77]. A partire da questa storica sentenza della CGUE, ogni cittadino di uno Stato membro può appellarsi al diritto all'oblio, chiedendo al fornitore di un motore di ricerca online o di un sito web l'eliminazione di determinate informazioni personali, anche attraverso l'eliminazione dei *link* verso pagine web dall'elenco dei risultati che appaiono dopo una ricerca a partire dal suo nome[78].

La tutela del diritto di ogni individuo a essere dimenticato, alla cancellazione dal web delle informazioni che lo riguardano, è continuamente messa a rischio dall'affiorare nello scenario globale di nuove tecnologie. In particolare, come accennato sopra, il *machine learning* si pone in tensione con questo diritto: dato che l'IA si serve dei dati degli utenti per accrescere le sue conoscenze, sorge il problema del rapporto di un tale utilizzo delle informazioni personali con il rispetto del principio personalistico e del diritto all'autodeterminazione dell'individuo, che rischia di non ottenere mai l'eliminazione delle sue informazioni personali dal circuito del web.

Affinché venga garantito il diritto all'oblio di ogni individuo anche quando le informazioni siano utilizzate per addestrare un'Intelligenza Artificiale, è stato collaudato lo strumento del *machine unlearning*. Non si tratta, come la sua denominazione potrebbe portare a ritenere, dell'opposto del *machine learning*: non è un processo inverso in cui un certo compito, che si è eseguito grazie all'uso di determinati dati, viene cancellato e dimenticato dall'IA. Il meccanismo che interviene consiste, invece, nel fare in modo che la macchina dimentichi come alcuni dati contribuiscono a creare un certo modello di *machine learning*, e ha

[77] Alagna, op. cit., p. 9.
[78] Alagna, op. cit., p. 18.

l'obiettivo quindi di annullare l'influenza di alcuni dati su un modello, garantendo quindi la cancellazione degli stessi[79]. Sebbene alcuni abbiano sostenuto[80] che il *machine unlearning* sia uno strumento promettente per assicurare il rispetto dei cinque principi per un'IA affidabile elaborati dall'OCSE[81], altri ritengono che tale strumento non sia l'unico efficace, e che comunque non sia sempre sufficiente[82].

Ad ogni modo, non privo di interesse è il circuito virtuoso che si crea nel rapporto tra i modelli di *machine unlearning* e la garanzia di esplicabilità. L'esplicabilità, oltre ad essere un requisito fondamentale per una buona IA, è anche un elemento facilitatore per la creazione di algoritmi di *machine unlearning*. Parrebbe infatti difficile sottrarre dal modello dei dati senza che siano, innanzitutto, conosciute le modalità con cui il modello stesso è stato costruito e attraverso le quali funziona. D'altra parte, anche lo stesso *machine unlearning* contribuisce a rendere i processi di IA più chiari e "spiegabili". Ciò poiché l'eliminazione di alcuni dati dal modello permette di osservare l'evoluzione dello stesso da prima a dopo la cancellazione degli stessi, agevolando una più ampia e completa comprensione dell'algoritmo[83].

2.2. Profili di criticità con riferimento al principio di eguaglianza

Un ulteriore ambito di interesse ai fini di questa trattazione è il rapporto tra l'Intelligenza Artificiale e il principio di eguaglianza. L'opacità che caratterizza i processi decisionali dell'IA infatti porta con sé dei rischi di discriminazione nei confronti di coloro che sono coinvolti nelle determinazioni algoritmiche, quando la non esplicabilità del meccanismo algoritmico si sostanzia nella non conoscibilità della motivazione su cui si fonda la decisione adottata. Ciò implica

---

[79] E. Hine, C. Novelli, M. Taddeo, L. Floridi, *Supporting Trustworthy AI Through Machine Unlearning*, Science and Engineering Ethics, National Library of Medicine, Springer 2024, p. 1.
[80] Ibid.
[81] Si tratta dei principi elaborati dal Consiglio dell'Organizzazione per la Cooperazione e lo Sviluppo Economico nella *Raccomandazione del Consiglio sull'Intelligenza Artificiale* adottata nel 2019, e aggiornata nel 2024 per affrontare i problemi posti dai nuovi sviluppi tecnologici, tra cui l'IA generativa.
[82] J. Xu, Z. Wu, C. Wang, X. Jia, *Machine Unlearnig: Solutions and Challenges*, pubblicato in IEE Transactions on Emerging Topics in Computational Intelligence, in www.semanticscholar.org (consultato il 28 Maggio 2026).
[83] J. Xu, Z. Wu, C. Wang, X. Jia, op. cit., p. 15.

un rischio per il rispetto del principio di trasparenza, - di cui si parlerà più estesamente in seguito -, e del principio di non discriminazione, poiché i soggetti interessati potrebbero non godere di tutele adeguate quale rimedio per eventuali illegittimità delle decisioni algoritmiche[84].

Il principio di eguaglianza è un campo di prova particolarmente interessante per osservare le problematiche legate alla presenza di *bias*. Il termine *bias* ha più significati, con accezione diversa a seconda che ad impiegarlo siano fisici o ingegneri elettronici che si occupano di IA; si tratta, in linea generale, di un "errore sistematico", un pregiudizio che può essere preesistente o introdotto successivamente. Sul piano giuridico è tacciabile di illegittimità, in quanto porta a trattamenti discriminatori per via di un'erronea comprensione della realtà. Tale rischio è particolarmente concreto nei casi di algoritmi a scopo predittivo, di cui si tratterà in seguito[85].

Sono individuabili due fonti principali di *bias* algoritmico: la prima riguarda la scelta dei dati, la seconda la rappresentazione dello stato del mondo. Nel primo caso, i dati possono essere poco rappresentativi: ad esempio, l'algoritmo individua erroneamente come consueto un legame tra grandezze che è invece secondario, o casuale, o ancora trae una legge generale da un fenomeno osservabile solo in una sottocategoria della popolazione, e non nell'insieme. Nel secondo caso invece, lo *status quo* del mondo in base al quale la decisione è assunta è espressivo di disuguaglianze che si riflettono nei dati. Gli effetti sono, da un lato, la creazione di una condizione deteriore per alcune categorie di persone, dall'altro la riproduzione di iniquità già presenti, che si cristallizzano e si rafforzano[86].

---

[84] M. Fasan, *Intelligenza artificiale e costituzionalismo contemporaneo: principi, diritti e modelli in prospettiva comparata*, Collana della Facoltà di Giurisprudenza dell'Università di Trento, 49, Università di Trento 2024, p. 95.

[85] E. Falletti, *Discriminazione algoritmica: una prospettiva comparata*, G. Giappichelli Editore, Torino 2022, p. 160.

[86] G. D'Acquisto, *Decisioni algoritmiche: equità, causalità, trasparenza*, G. Giappichelli Editore, Torino 2022, p. 45.

Il tema della "discriminazione algoritmica" è particolarmente centrale nel *Blueprint for an AI Bill of Rights*[87] statunitense. Si tratta di un documento non vincolante, ma particolarmente significativo per l'approccio attento alle "ricadute umane collettive"[88], e quindi all'impatto dell'IA sulla comunità. Il *Blueprint* si concentra sull'individuare principi e raccomandazioni che assicurino la conformità dei sistemi di Intelligenza Artificiale al principio di eguaglianza; e individua, in particolare, nel diritto di opzione una garanzia esercitabile dai cittadini dinanzi a scelte della macchina che siano state assunte senza il consenso della persona[89]. La tensione tra l'IA e il principio di eguaglianza, e quindi il problema della discriminazione algoritmica, è uno dei nodi centrali che il documento si propone di affrontare, mostrando così una certa lungimiranza sulle questioni che la società attuale e quella futura dovranno affrontare.

Questa dimensione collettiva della discriminazione artificiale è invece trascurata dall'*AI Act* dell'Unione Europea, che non pare avere una tale consapevolezza degli esiti che i sistemi di *machine learning* hanno sulla comunità, ancor prima che sul singolo: il Regolamento unionale fa infatti riferimento ai principi della Carta dei diritti fondamentali dell'Unione Europea e ad una concezione individualistica del principio di eguaglianza e di non discriminazione. Si tratta di un approccio distante da quello statunitense, il quale pone maggiormente l'accento sugli effetti dannosi che le decisioni algoritmiche discriminatorie hanno su individui e gruppi, intuendo a ragione quanto le violazioni del principio di eguaglianza in materia di IA si concretizzino sovente in discriminazioni collettive[90]. In ogni caso, nel diritto unionale e in quello nazionale, il principio di non discriminazione algoritmica non era previsto espressamente fino all'emanazione dell'*AI Act*; esso era contenuto unicamente nel considerando n.

---

[87] Si fa riferimento al *Blueprint for an AI Bill of Rights. Making Automated-systems work for the american people,* documento adottato dall'amministrazione federale degli Stati Uniti nell'ottobre del 2022.

[88] C. Nardocci, *Dalla "self-regulation" alla frammentata regolamentazione dei sistemi di intelligenza artificiale: uno sguardo alla (diversa) prospettiva statunitense*, "Diritto pubblico comparato ed europeo", Fascicolo 4, ottobre-dicembre 2024, in "Il Mulino" – Rivisteweb, p. 863.

[89] Ibid.

[90] Nardocci, op. cit., pp. 862-866.

71 del Regolamento 679/2016, ma non era riprodotto nel testo. Esso affermava che un modo per evitare la discriminazione fosse la rettifica dei dati di addestramento in "ingresso" [91].

In tal senso è utile richiamare un principio cui fanno riferimento i *data scientists*, che ben illustra come si manifestano le violazioni del principio di non discriminazione nell'ambito del *machine learning*. È noto come GIGO - "*garbage in garbage out*" -, e fa riferimento alla circostanza per cui un algoritmo riflette necessariamente la qualità dei dati su cui è costruito[92]. Le determinazioni algoritmiche infatti sono di per sé inficiate da *bias* ogni qual volta i dati di input che hanno addestrato l'IA sono discriminatori. La questione risiede dunque nella presenza di algoritmi intrinsecamente incostituzionali, per via dei dati su cui sono costruiti. La decisione artificiale viene fatta discendere, insomma, dalla realtà, che spesso è iniqua; al contrario, dovrebbe essere la determinazione a "normare" il reale secondo giustizia.

Su questo punto, appare opportuno citare il dibattito che si è sviluppato negli Stati Uniti a seguito della decisione *Compas*, di cui si parlerà più estesamente in seguito. Il rischio di recidiva predetto da questo sistema, su un campione di 10.000 imputati in Florida, era evidentemente sbilanciato in senso discriminatorio nei confronti degli imputati neri, con una sovrastima della probabilità di recidiva di questi ultimi rispetto agli imputati bianchi. Il giudizio sulla pericolosità del soggetto, e nello specifico sulla probabilità che compisse un altro reato nei due anni successivi, è risultato – sulla base della comparazione con il tasso di recidiva effettivamente realizzato dallo stesso campione di imputati in quel periodo di tempo – viziato da una discriminazione di tipo razziale, operata dall'algoritmo[93].

Quando si fa riferimento al principio di eguaglianza in relazione ai sistemi di Intelligenza Artificiale, ci si riferisce, tra le altre questioni, al *black box effect*. Si tratta del "problema dell'impossibilità, da parte degli stessi programmatori e tecnici, di spiegare […] come il sistema sia arrivato a determinati risultati o di

---

[91] A. Simoncini, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in "BioLaw Journal - Rivista di BioDiritto", 1/19, Febbraio 2019, Trento, p. 84.
[92] Simoncini, op. cit., p. 85.
[93] Simoncini, op. cit., pp. 85-86.

prevedere gli stessi risultati"[94]. Si è detto che il funzionamento degli algoritmi decisori utilizzati è spesso inconoscibile; ciò implica il rischio che l'erroneità, l'incompletezza o il carattere discriminatorio dei dati rimangano nascosti. I problemi più rilevanti in relazione alla tenuta del principio di non discriminazione sopraggiungono quando le decisioni algoritmiche vengono adottate attraverso modelli di *deep learning* realizzati con reti neurali artificiali. Queste ultime, che imitano l'attività cerebrale, immagazzinano i dati dall'esperienza e li elaborano con procedimenti nascosti, non visibili, cioè le sopracitate *black box*.

Una soluzione potrebbe essere la creazione di registri di dati, per far sì che questi, raccolti al fine di addestrare il sistema, generino delle determinazioni algoritmiche fondate su dati verificati e soggetti a controllo e sorveglianza umana. La conoscibilità dei dati e la garanzia sulla loro affidabilità, inoltre, porta con sé l'esplicabilità del processo decisionale, che può essere quindi sottoposto al vaglio del diritto: la predisposizione di tali registri ha la funzione non solo, quindi, di scongiurare questioni di illegittimità costituzionale con riferimento al principio di eguaglianza, ma anche in relazione al principio di trasparenza, per cui appare una delle soluzioni attualmente più convincenti. Ulteriore strumento per impedire che il potere decisionale della macchina raggiunga un livello tale di autonomia da non poter essere riassunto dall'uomo, è quello di prevedere meccanismi atti a garantire il ripristino del controllo, come l'alterazione o la disattivazione dei sistemi di *machine learning*[95].

È chiaro, in base a ciò che si è detto fino a questo punto, che gli algoritmi siano ben lontani dall'essere neutri, ma al contrario sono da considerare "opinioni umane strutturate in forma matematica che, come tali, riflettono, in misura più o meno rilevante, le precomprensioni di chi li progetta, rischiando di volgere la discriminazione algoritmica in discriminazione sociale"[96].

---

[94] S. Amato, *Inquietudini digitali. Il "black box effect"*, "Rivista di filosofia del diritto", Fascicolo 1, giugno 2025, in "Il Mulino – Rivisteweb", Bologna, p. 34.
[95] Falletti, op. cit., pp. 137-138.
[96] A. Soro, *Proteggere i dati per governare la complessità*, in www.garanteprivacy.it (consultato l'8 Giugno 2026).

Il dibattito che ha impegnato la dottrina circa la possibilità che un algoritmo di intelligenza artificiale possa avere, in questo senso, soggettività giuridica e quindi possa essergli imputata un qualche tipo di responsabilità, si è articolato in due direttrici in particolare: la brevettabilità degli algoritmi e il riconoscimento agli stessi algoritmi di diritti di proprietà intellettuale.

Quest'ultima questione, se risolta positivamente, porterebbe con sé un'ulteriore domanda rilevante ai nostri fini: se si possa parlare di autonomia decisionale automatizzata e, in caso di risposta positiva, a chi debba essere attribuita la responsabilità per gli effetti conseguenti alle decisioni discriminatorie adottate dall'IA. Possiamo sin da subito chiarire che, alla luce della disciplina attuale, è sempre il proprietario del sistema di IA ad essere titolare dei diritti di proprietà intellettuale, e di conseguenza la responsabilità per le decisioni da questa elaborate ricade sempre sulla persona fisica.

Negli Stati Uniti, la giurisprudenza è stata chiamata a pronunciarsi proprio su tale questione, e si è espressa nel senso di negare il riconoscimento dei diritti brevettuali su due invenzioni del sistema di Intelligenza Artificiale "DABUS" ("*Device for the Autonomous Bootstrapping of Unified Sentience*"), che – secondo quanto sostenuto dal creatore Stephen Thaler, attore delle istanze - possedeva i caratteri propri dell'"inventore" sulla base dei criteri disposti dalla normativa statunitense, in quanto "macchina creativa". Trattandosi di una *global litigation*, le domande di brevetto sono state oggetto di numerose decisioni giudiziarie, che nella quasi totalità hanno portato al rigetto dell'istanza. In particolare, le decisioni più rilevanti sono quella del *US District Court for the Eastern District of Virginia* e quella del *US Patent and Trademarks Office (USPTO).* Nel primo caso, la Corte ha rigettato l'istanza statuendo che la natura di macchina basata su un algoritmo di intelligenza artificiale rendeva impossibile l'attribuzione di una qualsivoglia titolarità di brevetti; la seconda, riconoscendo il valore politico delle istanze di Thaler, ha chiarito che solo il Congresso può rimodulare la definizione di "inventore" contenuta nella legge sui brevetti, e che poiché il riferimento alla persona fisica è evidente nel testo di legge, non può

esservi alcun riconoscimento di soggettività giuridica degli algoritmi per via giurisprudenziale[97].

È ormai fuor di dubbio, dunque, che l'Intelligenza Artificiale abbia una potenzialità discriminatoria; la questione risiede tutt'al più nelle modalità con cui tale discriminazione ha luogo, nei confronti di chi, e sulla base di quali criteri. In particolare, studiare le modalità con cui si verifica l'*AI-derived discrimination* è prodromico per comprendere se il diritto antidiscriminatorio classico, quello cioè delle Costituzioni e dei trattati di diritto internazionale, sia sufficiente ad arginare il fenomeno, ovvero si rendano necessarie categorie ulteriori rispetto a quelle tradizionali, atte a fronteggiare i fattori di discriminazione che riguardano specificatamente i sistemi di *machine learning*[98]. Un'ipotesi di particolare rilievo, poiché in essa emergono i tratti più caratteristici e distintivi della discriminazione causata da sistemi di IA, è la c.d. "*proxy discrimination*".

Si tratta di una peculiare fattispecie di discriminazione, definita "per associazione"[99]: in questi casi, la disparità di trattamento non si fonda su qualità individuali che siano immediatamente predittive dell'appartenenza a gruppi protetti, - non cioè su quelle qualità che tipicamente sono associate a fenomeni discriminatori, come l'etnia, il sesso, il genere, - ma su dati che sono associati a categorie protette in modo indiretto. In altre parole, il *proxy* può essere definito come un qualsivoglia elemento, - che sia descrittivo della fisionomia umana, ambientale, ovvero espressivo di una preferenza, - che presenti un'associazione con una caratteristica propria di una categoria protetta. L'uso di questo strumento deve essere sottoposto all'attenzione del giurista ogni qualvolta venga impiegato con finalità predittive: tutte le volte, quindi, in cui vi sia nel *proxy* la capacità di istituire correlazioni con fattori che generano trattamenti discriminatori, poiché è questo aspetto a generare una violazione del principio di eguaglianza.

Sebbene il *proxy* non rinvii a caratteristiche protette può di fatto generare un esito discriminatorio, attraverso associazioni operate tra elementi "neutri" e l'appartenenza a una certa categoria discriminata. Si tratta, nello specifico, della

---

[97] Falletti, op. cit., pp. 160-164.
[98] C. Nardocci, *Algoritmi, eguaglianza, discriminazione: le sfide dell'intelligenza artificiale*, G. Giappichelli Editore, Torino 2025, p. 62-65.
[99] Nardocci, op. cit., p. 102.

"*indirect proxy discrimination*": un esempio[100] di questa fattispecie discriminatoria è il caso di un algoritmo che abbia l'obiettivo di selezionare i candidati per una posizione lavorativa in cui è richiesto come requisito necessario l'altezza. Il *data-set* dell'algoritmo non possiede i dati né sul sesso dei candidati né sulla loro altezza, per via della correlazione tra altezza e sesso; tuttavia la macchina potrà comunque selezionare il candidato migliore – quindi il candidato che meglio soddisfi il requisito ricercato, cioè quello dell'altezza,- facendo ricorso ad un *proxy* che, pur non essendo direttamente un fattore di discriminazione, possa comunque essere indicativo del sesso di appartenenza dei candidati, ad esempio le preferenze di serie tv su una certa piattaforma streaming. Si vede come il ricorso ad un dato non sensibile e non vietato, apparentemente neutro e non direttamente indicativo di una categoria protetta, produca comunque un effetto discriminatorio, poiché rivelatore dell'appartenenza di alcuni candidati a una certa categoria discriminata, in questo caso il genere femminile[101].

Alla luce di quanto detto, la piena sovrapponibilità dei fattori di discriminazione tradizionali e il *proxy* è da escludere, in quanto, si è detto, esso non è di per sé discriminatorio; tuttavia il giudice è chiamato a sindacare sulla legittimità del suo impiego ogni qualvolta il *proxy* crei connessioni con altri dati qualificati come "sospetti" e tipizzati come discriminatori.

Si accoglie dunque in questa sede la tesi secondo la quale il carattere predittivo del *proxy*, se provato, rende lo strumento equiparabile ad un fattore di discriminazione tradizionale, tacciabile dunque di illegittimità costituzionale, con la conseguente attivazione delle tutele per i soggetti lesi[102]. In tal senso, il diritto antidiscriminatorio classico si presenta ad oggi ancora lo strumento più efficace per fronteggiare le sfide della discriminazione artificiale, a condizione che le sue categorie vengano interpretate alla luce delle peculiarità dell'IA, innanzitutto con la consapevolezza che i rischi insiti nei sistemi di *machine*

---

[100] Si veda A.E.R. Prince, D. Schwarcz, *Proxy discrimination in the age of artificial intelligence and big data*, in Iowa Law Review, 2020, p. 1280.
[101] Nardocci, op. cit., pp. 114-115.
[102] Nardocci, op. cit., pp. 198-199.

learning sono spesso di non facile individuazione, per via delle associazioni occulte che l'algoritmo opera tra una quantità enorme di dati.

Fino a questo momento, si è fatto riferimento ai casi in cui la discriminazione algoritmica si sia già verificata, e si prospetti quindi la necessità di un intervento *ex post* da parte del diritto, in particolare attraverso il sindacato del giudice. Tuttavia è possibile valutare preventivamente la bontà di un algoritmo, ad esempio con l'ausilio di modelli matematici, oppure attraverso il ripristino dell'equità tramite un processo chiamato "*debiasing*". L'espressione può essere tradotta con "eliminazione della distorsione", cioè del *bias*: si può ottenere attraverso "un criterio decisionale a soglia differenziato"[103], che si basi sull'applicazione di una soglia diversa per la categoria protetta rispetto alla categoria non protetta. L'introduzione di questo criterio realizza l'obiettivo di ottenere un eguale valore di parità demografica – cioè la stessa quota di giudicati positivi e di giudicati negativi per entrambe le categorie, - e una parità di opportunità più sostanziale – cioè la stessa percentuale di *true positive* e *false positive*[104].

È stato sostenuto[105] che questo processo sacrifichi il principio di parità di trattamento, sebbene si ritenga necessario rilevare in questa sede che, se attuato con le modalità sopra descritte, il *debiasing* è conforme al principio di eguaglianza così come enunciato all'art. 3 della Costituzione. Come detto nel capitolo precedente infatti, l'eguaglianza deve essere garantita non solo in senso formale, e in questo senso il *debiasing* si propone di garantire una parità di trattamento che sia sostanziale, effettiva. Ciò deve avvenire innanzitutto attraverso una presa d'atto, da parte del decisore che realizza l'algoritmo, delle iniquità insite nella società e, alla luce dello stato delle cose, deve avvenire una ricalibrazione dell'algoritmo stesso, che non è dunque da ritenere una violazione della parità di trattamento, ma al contrario un atto necessario affinché i destinatari della decisione algoritmica siano valutati su un piano di parità. Si vede dunque come l'interpretazione dei principi costituzionali in relazione al

[103] D'Acquisto, op. cit., p. 67.
[104] Per un approfondimento su questi concetti, si veda D'Acquisto, op. cit., pp. 46 e ss.
[105] D'Acquisto, op. cit., p. 67.

funzionamento dei nuovi sistemi di IA sia un elemento cruciale e quantomai attuale, di cui l'operatore del diritto deve essere consapevole per affrontare le questioni poste dalle nuove tecnologie.

2.3. Profili di criticità con riferimento al principio di trasparenza

L'avvento dell'Intelligenza Artificiale ha avuto effetti dirompenti in molti ambiti del diritto, a seguito del suo sempre più frequente impiego: tra i soggetti che fanno uso di questi sistemi vi è la pubblica amministrazione, che se ne serve per lo svolgimento dell'attività di regolazione economica, di erogazione dei servizi pubblici e di adozione delle decisioni[106].

Il tema che si intende affrontare in questa sede attiene l'assunzione della decisione amministrativa attraverso l'ausilio di algoritmi, e il suo rapporto con il principio di trasparenza. Ci si chiede in particolare se, e nel caso a quali condizioni, la pubblica amministrazione possa delegare l'esercizio della propria attività discrezionale a un sistema di *machine learning*. Occorre chiarire sin dal principio che, sebbene sia un tema complesso e disseminato di rischi e difficoltà interpretative, l'impiego di algoritmi per l'assolvimento di tale funzione ha numerosi vantaggi: permetterebbe innanzitutto di diminuire il rischio di errori umani e di ridurre i costi e i tempi dell'amministrazione stessa; inoltre avrebbe il pregio di aumentare enormemente il numero di dati raccolti e analizzati[107].

Una prima distinzione che si deve operare riguarda la differenza tra algoritmi deterministici e non deterministici, e le rispettive potenzialità di impiego. Negli algoritmi deterministici *l'input* che viene dato alla macchina è la disposizione costituzionale, legislativa o regolamentare, mentre l'*output* è predeterminato dalla disposizione stessa, ed è dunque assolutamente prevedibile: per questo possono essere impiegati dalla pubblica amministrazione solo nell'esercizio di attività vincolata. Gli algoritmi non deterministici invece sono capaci di autoapprendimento, supervisionato o non supervisionato: hanno un enorme

---

[106] B. Marchetti, *La garanzia dello* human in the loop *alla prova della decisione amministrativa algoritmica*, in BioLaw Journal – Rivista di Biodiritto, 2/2021, p. 368.

[107] G. C. di San Luca, M. Paladino, *Decisione amministrativa e intelligenza artificiale,* in federalismi.it (consultato il 15 Giugno 2026), 29 gennaio 2025. !

livello di autonomia nella selezione dei dati rilevanti e nella ricerca delle possibili correlazioni, che vengono estrapolati dai c.d. *data lakes*.

Questi ultimi possono essere definiti come il luogo in cui vengono archiviati, analizzati e correlati tra loro i dati di varia tipologia e provenienza[108]. Negli algoritmi non deterministici l'*input* è dato dall'analisi autonoma di decisioni precedenti da parte della macchina; quanto all'*output*, che consiste nella decisione amministrativa discrezionale, è elaborato dall'algoritmo sulla base della similitudine o della contrapposizione con le decisioni relative a vicende pregresse dell'amministrazione[109]. Essi potrebbero essere adoperati per l'esercizio di attività amministrativa discrezionale da parte dell'amministrazione, tuttavia le posizioni di dottrina e giurisprudenza sono tutt'altro che concordi.

La giurisprudenza italiana, ancor prima dell'entrata in vigore dell'*AI Act*, si è espressa nel senso di una precisa limitazione dell'autonomia della macchina nell'adozione delle decisioni amministrative discrezionali, discostandosi e superando la disciplina contenuta all'art. 22 del GDPR: il giudice amministrativo ha accordato una garanzia più ampia a protezione della partecipazione umana nei processi decisionali. Il testo del regolamento europeo, nell'imporre il divieto di decisioni esclusivamente automatizzate, contempla infatti tre eccezioni talmente ampie da vanificare la garanzia stessa. Il paragrafo 2 dell'articolo individua le ipotesi in cui il divieto non si applica nei casi in cui: "*a) sia necessaria per la conclusione o l'esecuzione di un contratto tra l'interessato e un titolare del trattamento; b) sia autorizzata dal diritto dell'Unione o dello Stato membro cui è soggetto il titolare del trattamento, che precisa altresì misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato; c) si basi sul consenso esplicito dell'interessato*[110]". Il diritto unionale appare

---

[108] G. Avanzini, *Intelligenza artificiale,* machine learning *e istruttoria procedimentale: vantaggi, limiti ed esigenze di una specifica* data governance, in A. Pajno, F. Donati, A. Perrucci (a cura di), Intelligenza artificiale e diritto: una rivoluzione?, Volume 2, Amministrazione, responsabilità, giurisdizione, Asrtrid, Il Mulino, Bologna 2022, p. 80.

[109] R. Cavallo Perin, *Ragionando come se la digitalizzazione fosse data*, in Diritto amministrativo, Rivista trimestrale, 2/2020, p. 310.

[110] *Regolamento (UE) 2016/679*, in Gazzetta Ufficiale dell'Unione Europea, art. 22 paragrafo 2.

dunque molto più propenso di quanto possa sembrare all'impiego esclusivo di decisioni automatizzate.

La giurisprudenza amministrativa italiana ha invece affermato che una decisione algoritmica completamente automatizzata è di per sé incostituzionale, in quanto in contrasto con gli artt. 3, 24 e 97 della Costituzione e con l'art. 6 della Convenzione europea dei diritti dell'uomo: sul piano nazionale, dunque, la tutela accordata è molto più intensa di quella prevista dal GDPR. Emergono in questa materia le differenti modalità di bilanciamento tra, da un lato, la tutela della dignità della persona in ambito digitale – quella che viene definita la *digital persona,* - e dall'altra la garanzia della libera circolazione dei dati, quale risorsa economica fondamentale. L'Unione Europea, nel bilanciare questi due opposti interessi, individua il punto equilibrio in modo più favorevole alla libertà di circolazione dei dati, mentre a livello statale prevale la tutela della libertà della persona[111].

Mentre è assolutamente condivisibile l'idea che la completa attribuzione del potere decisione all'algoritmo sia incostituzionale, non pare convincente l'orientamento di una parte della giurisprudenza amministrativa di primo grado per cui l'automatizzazione dell'attività amministrativa andrebbe esclusa *in toto*, perché in contrasto con il principio di trasparenza e con la disciplina sulla partecipazione procedimentale, ma anche con l'obbligo di motivazione dei provvedimenti amministrativi ex art. 3, co. 1, L. 241/1990. Secondo tale orientamento, l'adozione di decisioni da parte della macchina contravverrebbe "*al principio dell'interlocuzione personale intessuto nell'art. 6 della legge sul procedimento e a quello ad esso presupposto di istituzione della figura del responsabile del procedimento*", e non sarebbe in grado di sostituire "*l'attività cognitiva, acquisitiva e di giudizio che solo un'istruttoria affidata ad un funzionario persona fisica è in grado di svolgere*" e quindi deve essere il responsabile del procedimento a dominare "*le stesse procedure informatiche predisposte in funzione servente e alle quali va dunque riservato tutt'oggi un*

[111] A. Simoncini, *Amministrazione digitale algoritmica. Il quadro costituzionale*, in R. Cavallo Perin, D. U. Galetta (a cura di), Il diritto dell'amministrazione pubblica digitale, Giappichelli Editore, Torino 2020, p. 29.

ruolo strumentale e meramente ausiliario in seno al procedimento amministrativo e giammai dominante o surrogatorio dell'attività dell'uomo[112]".

Il Consiglio di Stato, invece, si è espresso nel senso dell'ammissibilità dell'impiego di algoritmi informatici nell'adozione del provvedimento amministrativo, prima per i soli casi di attività vincolata, e successivamente anche per le ipotesi di attività discrezionale, ma con le opportune cautele. Il giudice chiarisce infatti che l'impiego di algoritmi è ammissibile a condizione che siano assicurati alcuni principi di "legalità algoritmica": la conoscibilità e la comprensibilità del modulo impiegato, del procedimento usato per la sua elaborazione, del meccanismo decisorio, ivi comprese le priorità e i dati considerati rilevanti nella procedura valutativa e decisionale[113].

Si tratta di un caso di *ex post regulation*, in cui il giudice è stato chiamato a operare un bilanciamento tra i principi costituzionali e l'impiego dell'IA. Si vede come la giurisprudenza si prospetti, nelle contingenze dei continui progressi scientifici in materia di *machine learning*, lo strumento più adatto alla sua regolazione: trattandosi di un fenomeno recente, di difficile definizione ed estremamente mutevole, il solo impiego di fonti primarie o secondarie rischia di non essere efficace - posto che, in ogni caso, non pare prospettabile un'adeguata regolazione dell'Intelligenza Artificiale che non sia anche, e soprattutto, sovranazionale[114].

L'impiego degli algoritmi nell'ambito della pubblica amministrazione porta con sé una riflessione sul rapporto tra uomo e macchina. Al di là di atteggiamenti pessimistici che spesso conducono ad un ingiustificato rifiuto e ad una demonizzazione dell'IA, è opportuno riflettere su alcune considerazioni di principio. Le classificazioni distinguono quei sistemi di IA che possono svolgere le proprie funzioni in totale autonomia ("*human out of the loop*"), da quelli che sono completamente dipendenti dal controllo umano ("*human in command*"); vi è poi una terza opzione, che si caratterizza per una qualche forma di integrazione tra intelligenza umana e intelligenza della macchina. Ciò può avvenire attraverso

[112] Si veda T.A.R. Lazio-Roma, sez. III Bis, 10/9/2018, n. 9224.
[113] Si veda Cons. Stato, Sez. VI, 13/12/2019, n. 8472.
[114] B. Marchetti, op. cit., p. 370.

un controllo successivo sull'operato della macchina, o attraverso il coinvolgimento dell'essere umano nel processo di decisione algoritmica: si parla in tal senso di "*human in the loop*"[115].

La garanzia dello "*human in the loop*", è ciò che deve guidare legislatore e interprete, poiché è permette il rispetto dei principi del diritto amministrativo, - cioè la trasparenza, i diritti di partecipazione, il diritto alla motivazione della decisione amministrativa -, senza però rinunciare ai benefici che l'introduzione dell'IA apporta in termini di riduzione dei costi e dei tempi dell'amministrazione pubblica. Ci si muove, dunque, nell'ambito della necessità di dover definire il "ruolo che il funzionario pubblico può concretamente svolgere in termini di guardiano della macchina"[116]. Lo "*human in the loop*" è emerso di recente tra gli studiosi che si occupano del complesso rapporto tra l'Intelligenza Artificiale e l'essere umano. Con questa espressione si fa riferimento a una delle condizioni atte a garantire la c.d. "meta-autonomia", cioè lo sviluppo antropocentrico delle nuove tecnologie, al fine di scongiurare il rischio che l'uomo venga completamente sostituito dalla macchina, senza essere in grado di riassumerne il controllo. Alla luce di quanto detto nel corso di questa trattazione, si vede come tale risultato non sia certo di facile ottenimento, specialmente nel caso di sistemi di *machine learning*, che si caratterizzano per il fenomeno della *black box*, cioè della non conoscibilità e non prevedibilità dei processi decisionali, sconosciuti agli stessi programmatori.

Si prospetta in questo senso la possibilità che il funzionario pubblico, in quanto principale utilizzatore del sistema, debba essere selezionato anche sulla base delle proprie competenze matematiche ed informatiche, necessarie al fine di esercitare un controllo reale sulla macchina[117]. Il paradigma della "co-intelligenza"[118] subisce così un ulteriore ampliamento: questo termine, a cui ci si riferisce per indicare il coinvolgimento di conoscenze esterne all'amministrazione ad integrazione della stessa, oggi attiene anche

---

[115] C. Casonato, B. Marchetti, *Prime osservazioni sulla Proposta di Regolamento dell'Unione europea in materia di intelligenza artificiale*, in "BioLaw Journal – Rivista di BioDiritto", 2021, n. 3, p. 420.
[116] B. Marchetti, op. cit., p. 377.
[117] B. Marchetti, op.cit., pp. 371 - 378.
[118] Così definito in E. Mollick, *Co-intelligence*, Londra, WH Allen, 2024.

all'integrazione dell'intelligenza amministrativa attraverso l'operato dell'IA[119]. Ciò richiederà negli anni a venire una riflessione sul ruolo del funzionario pubblico nei contesti ad alta automazione: si pensi al fatto che un dipendente pubblico – che secondo le stime attuali impiega un terzo del proprio tempo lavorativo in operazioni ripetitive, quali in raccolta di informazioni -, sarà sostituito in molte di queste funzioni basilari dalla macchina.

Le politiche di reclutamento dovranno quindi riorientarsi nel senso di selezionare la forza lavoro sulla base di criteri differenti, ricercando competenze di tipo manageriale, strategico e tecnico, a discapito delle competenze generaliste, ad oggi molto richieste. Sarà poi necessaria un'attività di formazione costante, affinché le competenze di dipendenti pubblici e il progresso tecnologico si evolvano in simultanea, senza che la mancanza di aggiornamento nella formazione generi uno squilibrio che rischi di vanificare il ruolo di vigilante del funzionario. Anche la valutazione della prestazione professionale individuale dovrà cambiare i suoi parametri di riferimento: qualità come la capacità di elaborazione di pensiero critico e creativo, l'impegno all'apprendimento costante, le conoscenze digitali, saranno determinanti nella fase selettiva. Infine, le sempre più sofisticate forme di sorveglianza algoritmica richiedono di apprestare tutele salde per i dipendenti pubblici, sottoposti a controlli molto stringenti sulla loro attività lavorativa, come il monitoraggio sulla frequenza delle battute a tastiera, di invio della posta elettronica aziendale, di utilizzo di risorse comuni.

A tal proposito, negli Stati Uniti il disegno di legge introdotto nel 2024, lo *Stop Spying Bosses Act*, propone l'introduzione di un obbligo di trasparenza a carico dei datori di lavoro nei confronti dei dipendenti, sulle modalità di raccolta e trattamento dei dati relativi alle prestazioni professionali[120].

Alla luce di quanto detto sulle nuove esigenze che la pubblica amministrazione deve essere in grado di soddisfare per garantire il rispetto del principio dello *human in the loop*, appare evidente quanto siano importanti, e quanto quindi

---

[119] G. Sgueo, *La funzione pubblica sintetica. Tre domande su intelligenza artificiale generativa e pubblica amministrazione*, in Rivista Trimestrale di Diritto Pubblico, num. 4/2024, p. 1185.
[120] G. Sgueo, op. cit., pp. 1185-1190.

debbano essere meditate, le modalità di trasformazione della pubblica amministrazione nell'ottica di un adeguamento all'evoluzione tecnologica. La garanzia dello HITL, insomma, è facilmente professabile in linea di principio, ma molto meno semplice ad attuarsi, se si pensa alle odierne caratteristiche della pubblica amministrazione. Il c.d. "potere pubblico sintetico"[121], - con questa espressione si intende il potere pubblico che sempre di più si serve di sistemi informatici e algoritmi, - ha portato con sé la necessità che gli Stati individuassero strutture e organismi differenti con funzioni di assistenza, coordinamento o regolazione. La Commissione europea ha istituito a questo scopo l'Ufficio europeo per l'IA; in altri casi si è operata una scelta nel senso del potenziamento di strutture già esistenti, ad esempio l'Agenzia per l'Italia digitale; talvolta sono stati istituiti organismi informali di tipo consultivo, come è accaduto nel Regno Unito con il *Policy Lab*; ancora, Stati come Singapore, la Finlandia e gli Stati Uniti hanno optato per la stipulazione di accordi con aziende quali *Google*, *OpenAI* e *ChatGPT*[122]. La gamma di soluzioni prospettabili è vasta, a riprova del carattere sfidante dell'utilizzo dei sistemi di IA nella pubblica amministrazione.

I limiti e le esigenze legate all'ingresso degli algoritmi nel procedimento amministrativo, quindi, sono molteplici, e richiedono una specifica *data governance*. Ciò poiché i tradizionali principi di diritto amministrativo, a partire dal principio di trasparenza della pubblica amministrazione, sono ostacolati dall'operare degli algoritmi, sia nella fase istruttoria che in quella decisoria. I percorsi non sono sempre comprensibili, e gli errori difficilmente individuabili degli algoritmi, - sia nella fase di accertamento dei fatti e degli interessi che in quella della valutazione degli stessi, - rendono le tradizionali regole del diritto amministrativo inadeguate a far fronte alle peculiari caratteristiche e al diverso grado di autonomia e prevedibilità dei sistemi di IA. Viene chiesto al giurista, dunque, di declinare tali principi – completezza dell'istruttoria, logicità e ragionevolezza della decisione, - in modo da garantirne il rispetto nei casi in cui

[121] G. Sgueo, op. cit., p. 1190.
[122] Ibid.

l'istruttoria sia invisibile, la decisione simultanea, e l'intervento umano solo eventuale e comunque preventivo[123].

La scelta del sistema da utilizzare non è quindi da ritenere una decisione neutra, di carattere meramente amministrativo: è richiesta una legittimazione che garantisca il rispetto, nella scelta dei sistemi di IA da utilizzare, del principio di legalità. Si rende quindi necessaria un'analisi sui limiti del sistema che si vuole utilizzare, e tale operazione può consistere in una valutazione preventiva di impatto del sistema. Essa individua le criticità che il sistema può far emergere nel rapporto tra pubblica amministrazione e cittadino, anche attraverso il ricorso a tecniche di *data governance* e *data management* che effettuino preventivamente un controllo sulla raccolta di dati, ivi compresa l'adeguatezza e la qualità degli stessi, il grado di autonomia decisionale del sistema che si ritiene tollerabile, e gli errori in cui l'IA potrebbe incorrere, e che potrebbero inficiare l'esito finale[124].

2.4. Nuove prospettive interpretative

Nei paragrafi precedenti di questo capitolo si è tentato di individuare le principali criticità che i fenomeni legati allo sviluppo dell'Intelligenza Artificiale portano con sé, in particolare quelle questioni che coinvolgono la dimensione dei diritti fondamentali. Il costituzionalismo contemporaneo è chiamato a fronteggiare questa sfida attraverso il ripensamento delle proprie categorie tradizionali, alla luce delle nuove esigenze di tutela dei principi cardine delle società democratiche, tra i quali il principio personalistico, il principio di uguaglianza e il principio di trasparenza.

Sono da rifuggire, occorre precisarlo, quelle derive neoluddiste che vedono nell'innovazione tecnologica in generale, e nell'Intelligenza Artificiale in particolare, un pericolo indiscriminato, indipendentemente dalle modalità e dalle finalità del suo impiego. Al contrario, l'automazione algoritmica comporta numerosi benefici in molti campi della società: possiamo ricondurli a tre

[123] G. Avanzini, op. cit., pp. 75-76.
[124] G. Avanzini, op. cit., pp. 94-96.

principali e generici vantaggi, - già in parte individuati sopra, - di cui gli operatori del diritto devono essere consapevoli, al fine di scongiurare il rischio che un'eccessiva diffidenza porti a una mutilazione delle potenzialità dello strumento.

Innanzitutto, i sistemi di IA riducono in maniera significativa i tempi per l'adozione di decisioni e lo svolgimento di certe funzioni, grazie alla rapidità con cui vengono analizzate le informazioni prima, e vengono individuate le correlazioni rilevanti tra i dati esaminati poi. In secondo luogo, le soluzioni proposte dell'IA sono particolarmente efficaci, sia per la significativa riduzione dei margini di errore, sia per la capacità di personalizzazione dei contenuti proposti, che sempre più coincidono con le necessità e i desideri delle persone. Infine, l'impiego di questi sistemi comporta un notevole risparmio economico, per via della diminuzione dei costi di processi predittivi e decisionali che prima erano realizzati dagli esseri umani, e che ora sono spesso demandati alla macchina[125].

Si tratta, a ben vedere, di una vera e propria rivoluzione, che necessita però delle opportune cautele e riflessioni interpretative per evitare che il nuovo paradigma di società digitale, che si va delineando a partire dalla fine del XX secolo, porti con sé una ingiustificata compressione, o addirittura una negazione di fatto, dei diritti fondamentali. È possibile sintetizzare specularmente, dunque, tre principali criticità legate ad un uso incontrollato dei sistemi di IA. Nel corso della trattazione si è tentato infatti di individuare alcune delle più evidenti zone d'ombra dell'impiego di algoritmi, in relazione alle quali è necessaria una valutazione sulla tenuta dei principi costituzionali.

Innanzitutto, la capacità dei sistemi di IA di attingere a un enorme bacino di dati, i c.d. *data lakes*, e di elaborarli attraverso l'individuazione di moltissime correlazioni – molte più di quelle che potrebbero mai essere individuate dalla mente umana – tra gli stessi, comporta il rischio che, qualora i dati impiegati fossero "imprecisi, viziati, inaccurati e carenti"[126], venga sacrificata la stessa

---

[125] M. Fasan, *I principi costituzionali nella disciplina dell'Intelligenza Artificiale. Nuove prospettive interpretative*, in C. Casonato, M. Fasan, S. Penasa (a cura di), Diritto e Intelligenza Artificiale, Sezione monografica – DPCE ONLINE, Fascicolo n. 1/2022, pp. 184-185.
[126] Fasan, op. cit., p. 186.

bontà dei processi decisionali. In particolare, non solo si rischia che il risultato ottenuto sia di qualità inferiore, in termini di efficienza, rispetto a quello che sarebbe stato conseguito da un essere umano, ma vi è anche il pericolo ben più grave che i risultati ottenuti generino delle discriminazioni verso alcune categorie di soggetti. Nel corso del capitolo si è cercato di evidenziare quanto l'analisi e il controllo sui dataset sia fondamentale in questo senso: occorre effettuare tutte le verifiche necessarie affinché il bacino di dati da cui attinge il sistema non sia viziato e pregiudizievole. L'impiego dell'Intelligenza Artificiale non può prescindere perciò da un processo di rimozione dei *bias* presenti nei dati raccolti.

In secondo luogo, l'accesso ad una grande quantità di dati fa sì che questi sistemi abbiano enormi potenzialità manipolative, con il rischio di incursioni di interessi esterni e terzi in dimensioni decisionali private[127]. La personalizzazione dei contenuti, se da una parte permette che si realizzi in modo agevole l'incontro tra l'offerta del web e le necessità dell'utente, dall'altra rischia di influenzare, se non addirittura determinare, le scelte dell'utente stesso senza che questi ne abbia consapevolezza. Lo scenario che si profila è allarmante, se si considera che i sistemi di IA sono in mano a privati; le *Big Tech*, grazie ad una profilazione sempre più accurata dell'utente, non si limitano più ad intercettare i suoi desideri, ma sono in grado addirittura di prevederli, anticiparli, fino ad arrivare ad un tale condizionamento per cui scelte, gusti, opinioni, convinzioni politiche, sono in gran misura indotti dall'algoritmo. Occorre quindi che venga garantita la possibilità di negare il consenso circa l'utilizzo dei dati personali e, d'altra parte, venga riconosciuto il diritto ad essere informati su quale sia il tipo di dati di cui si servirà il sistema, e per quali finalità.

Un modo ulteriore per scongiurare il rischio di una deriva deterministica consiste nell'incremento dei finanziamenti pubblici nel settore dell'IA, per favorirne la conformità ai principi tutelati dall'ordinamento costituzionale, attraverso un controllo da parte delle istituzioni[128]. Attualmente ciò che emerge è

[127] Fasan, op. cit., p. 195.
[128] Fasan, op. cit., p. 197.

un'"egemonia computazionale"[129], cioè quel fenomeno per cui a gestire i sistemi di IA sono poche aziende in tutto il mondo, le uniche in grado di sostenere i costi per l'addestramento dei sistemi stessi. Ciò comporta degli squilibri innanzitutto nel rapporto tra Stati e privati: è sempre più evidente la disuguaglianza informativa a favore dei privati, che rende i tentativi di controllo su questi strumenti da parte dei governi spesso inefficaci. Per comprendere la marginalità che oggi il mondo istituzionale ed accademico ha nello sviluppo di sistemi di IA, basti pensare che nel 2022, dei più importanti modelli di intelligenza artificiale prodotti, trentadue sono stati prodotti dall'industria, mentre solo tre in ambito universitario[130].

A determinare queste asimmetrie, si ritiene, non è la mancanza di regolazione né la frammentarietà normativa. Innanzitutto, la regolamentazione dell'IA si caratterizza, al contrario, per un surplus di norme: oltre l'Unione Europea, centoventisette Paesi hanno introdotto provvedimenti normativi in materia; nel 2023 si contavano[131], e si può ritenere ragionevolmente che ad oggi il numero sia aumentato, ben milleseicento *policy* vigenti in materia a livello globale. Tra queste, le Strategie nazionali sull'IA, i codici di condotta delle organizzazioni professionali e di settore, i *memorandum* di intesa internazionali e le sperimentazioni controllate. In particolare, queste ultime sono state impiegate nel Regno Unito e in Spagna per permettere agli operatori privati di verificare la conformità dei modelli di IA alle disposizioni normative, attraverso l'utilizzo di dataset protetti. In secondo luogo, non si può imputare la non omogeneità dello sviluppo dell'IA ad una frammentarietà nella regolazione della materia, poiché ormai gli Stati paiono concordi nel ritenere che l'uniformità debba essere una caratteristica fondamentale. Il problema, dato che non riguarda né la quantità né la frammentarietà della normazione, deve riguardare a ben vedere la qualità della stessa, che accentua la questione dell'evoluzione dell'Intelligenza Artificiale a velocità multiple[132].

---

[129] Sgueo, op. cit., p. 1193.
[130] Ibid.
[131] Si tratta del censimento effettuato da Deloitte nel 2023, in particolare: Deloitte Center for Government Insights, *The AI regulations that aren't being talked about*, 2023.
[132] Sgueo, op. cit., p. 1194.

Infine, spesso accade che non sia possibile risalire al procedimento che ha portato all'adozione della decisione finale. Si tratta del problema della *explainability* dei sistemi di IA, cioè della loro intellegibilità. Occorre dunque che il legislatore, nazionale e sovranazionale, elabori delle normative volte a garantire la tracciabilità dei dati e dei procedimenti: tutte le informazioni su come vengono raccolti e classificati i dati, sul tipo di algoritmo usato e sui risultati ottenuti devono essere documentate, in modo che si possa operare una verifica sulla correttezza della decisione finale adottata dalla macchina[133]. Un principio che deve essere soddisfatto per una buona IA, dunque, è quello della trasparenza dei processi, prodromica alla possibilità di motivare le soluzioni automatizzate. Il fenomeno della *black box*, molto frequente nei sistemi di *machine learning*, può essere arginato attraverso l'elaborazione di modelli di IA interpretabili o con l'impiego di strumenti tecnici che consentano di avere spiegazioni sul comportamento del sistema. Inoltre, è necessario che i destinatari delle decisioni adottate siano informati della loro interazione con una macchina, e quindi che siano consapevoli delle capacità e dei limiti della stessa[134].

Con riferimento a quest'ultimo aspetto, e cioè alla consapevolezza degli individui sul funzionamento dei sistemi di IA e su come essi debbano essere utilizzati, è opportuno fare ancora una riflessione. L'accessibilità dell'IA è un tema che mette in gioco, ancora una volta, il principio di non discriminazione. È opportuno che il legislatore tenga in considerazione la necessità che l'utilizzo e la comprensione di questi sistemi debba essere garantito ad un numero molto ampio di persone, specialmente a quelle categorie vulnerabili che dall'utilizzo potrebbero trarne maggiori vantaggi[135].

In definitiva, nel valutare la tenuta della Costituzione rispetto alle nuove questioni legate alle evoluzioni dell'Intelligenza Artificiale, pare utile ricordare quanto autorevolmente sostenuto da Gustavo Zagrebelsky, ex giudice della Corte Costituzionale: i principi non si applicano, si interrogano e si adeguano alle

[133] Fasan, op. cit., p. 191.
[134] Fasan. op. cit., p. 193.
[135] Fasan, op. cit., p. 194.

esigenze del caso da regolare[136]. La giurisdizione diventa, dunque, casistica[137]; e i casi da regolare necessitano del riconoscimento di principi che si adeguino alle esigenze concrete.

L'adeguamento dei principi è un compito che legislatore e giudice devono svolgere insieme, posto che in ogni caso il legislatore ordinario può smentire quanto affermato dalla giurisprudenza dei giudici comuni. Accade però che, dinanzi alle omissioni del legislatore, la giurisprudenza sia chiamata a intervenire per colmare il vuoto normativo. L'ampiezza della discrezionalità delle decisioni dei giudici dipende dunque dall'attività del legislatore: maggiori sono le lacune, maggiore sarà la discrezionalità giurisprudenziale[138]. D'altra parte però il sorgere di tecnologie sempre più sofisticate rende la possibilità di una tempestiva risposta legislativa difficile da immaginare; sono i giudici, dunque, i primi chiamati ad operare il bilanciamento tra progresso tecnologico e principi costituzionali, affinché l'innovazione non prevalga sui diritti fondamentali, né questi ostacolino irragionevolmente il progresso.

[136] Zagrebelsky, op. cit., pp. 769-770.
[137] Ibid.
[138] Zagrebelsky, op. cit., p. 772.

Capitolo terzo

Una nuova dimensione per i diritti nel settore della giustizia: rischi e benefici

3.1. L'impiego degli algoritmi in ambito penale: sistemi di giustizia predittiva e di polizia predittiva

Con l'espressione giustizia predittiva si intende innanzitutto un sistema che, attraverso l'ausilio di algoritmi, è in grado di prevedere l'esito di una controversia in base alle soluzioni precedentemente adottate in casi analoghi[139]. La giustizia predittiva è quindi la giustizia prevedibile: non solo nel senso della prevedibilità della disposizione da applicare, ma anche della prevedibilità della decisione giudiziale che verrà adottata[140].

Gli algoritmi possono predire l'esito di un giudizio attraverso calcoli statistici, sulla base dei precedenti giudiziari adottati per casi con caratteristiche simili; le ultime innovazioni nell'ambito dell'IA generativa sono persino in grado di riprodurre il ragionamento giuridico di un giudice ed emanare una decisione. Ad eccezione delle ipotesi in cui, per via delle peculiarità del caso di specie che rendano impossibile schematizzarlo per ricondurlo a una precisa categoria, non sia possibile prevedere l'esito del processo sulla base dei precedenti, nella maggior parte dei casi è possibile replicare il processo argomentativo proprio delle decisioni giudiziarie, individuando i passaggi logici e traducendoli in linguaggio informatico[141].

Interessa fare una precisazione sul tipo di sistema di intelligenza artificiale che viene impiegato quando si tratta di giustizia predittiva: si tratta di intelligenze artificiali "deboli" o "moderate", che possono fornire prestazioni specifiche superiori a quelle umane unicamente per quantità, e non anche per qualità. Esse si distinguono dalle intelligenze artificiali "forti", le quali sono in grado di

---

[139] C. Benetazzo, *IA, giustizia e P.A.: la sfida etica ed antropologica*, in federalismi.it, 12 Marzo 2025 (consultato il 9 Luglio 2026), p. 8.

[140] L. Viola, *Interpretazione della legge con modelli matematici. Processo, a.d.r., giustizia predittiva*, Volume I, Seconda edizione, Centro Studi Diritto Avanzato Edizioni, Milano 2018, p. 168.

[141] A. Carboni, *Giustizia algoritmica e applicazione dei precedenti parlamentari*, in D. De Lungo, G. Rizzoni (a cura di), Le assemblee rappresentative nell'era dell'intelligenza artificiale: profili costituzionali, G. Giappichelli Editore, Torino 2025, p. 225.

risolvere autonomamente problemi specialistici molto diversi tra loro, e forniscono prestazioni superiori a quelle umane anche per qualità.

In particolare, nell'analizzare la giurisprudenza precedente per prevedere un esito giudiziale, l'IA compie due attività: innanzitutto, il *natural language processing* e in secondo luogo il *machine learning*. Il sistema, dopo aver adottato una tecnica di trattamento del linguaggio umano, elabora dei modelli matematici atti a prevedere una futura decisione giudiziaria. La costruzione di questi modelli avviene attraverso la ricerca autonoma di correlazioni tra un'enorme mole di dati, in particolare i parametri di decisioni giudiziarie emanate in passato su un certo argomento. Si pensi, ad esempio, al caso in cui un sistema di giustizia predittiva sia chiamato a prevedere l'ammontare dell'assegno di mantenimento stabilito dal giudice all'esito di un divorzio giudiziale: nel costruire il modello matematico necessario a tal fine, la macchina dovrà individuare le correlazioni tra una serie di parametri che sono stati valutati come rilevanti dalla giurisprudenza precedente in materia, come la durata del matrimonio, il reddito dei coniugi, o ancora il verificarsi di un adulterio. La giustizia predittiva non consiste, dunque, - come suggerirebbe l'espressione, - nella "predizione" di eventi futuri, ma nella "previsione" degli stessi, intesa come il risultato dell'osservazione dei dati[142]. Con l'utilizzo delle tecniche sopra descritte, il grado di esattezza della macchina nel prevedere le decisioni di giudici è del 79%, secondo uno studio compiuto su 584 decisioni di apprendimento automatico della Corte europea dei diritti dell'uomo[143].

L'impiego di algoritmi nel settore della giustizia ha il merito di conferire maggiore certezza al diritto e garantire la prevedibilità delle decisioni, come accade nel sistema penale cinese: ogni qualvolta una decisione giurisdizionale si

---

[142] C. Barbaro, *Uso dell'intelligenza artificiale nei sistemi giudiziari: verso la definizione di principi etici condivisi a livello europeo? I lavori in corso alla Commissione europea per l'efficacia della giustizia (Cepej) del Consiglio d'Europa*, Obiettivo 1. Una giustizia (im)prevedibile, in "Questione Giustizia", n. 4/2018, Magistratura Democratica 2018 (consultato l'11 Luglio 2026), pp. 189-191.

[143] Si tratta di uno studio condotto dallo University College of London, cui la CEPEJ ha fatto riferimento nella *Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari* del 2018.

discosti dall'orientamento giurisprudenziale consolidato, viene generato un *warning* che segnala la difformità[144].

Il valore sociale della prevedibilità delle pronunce come effetto positivo dell'utilizzo dell'IA in ambito giudiziario è stato riconosciuto anche da una parte della giurisprudenza italiana. Da un lato la prevedibilità della sentenza porta dei benefici ancor prima dell'instaurazione di un giudizio, per la sua funzione deterrente avverso le domande giudiziali pretestuose, in quanto sarebbe possibile per l'attore valutare in anticipo la probabilità del verificarsi di un certo esito. Dall'altro, nella fase propriamente decisoria, i giudici che con la loro decisione si discostino dalla giurisprudenza che li ha preceduti, avranno maggiore contezza di questa loro operazione[145].

La giustizia predittiva porta però con sé dei rischi: ciò è stato chiarito anche dalla Commissione per l'efficienza della giustizia del Consiglio d'Europa (CEPEJ), che già nel 2016 si occupò di giustizia predittiva, compiendo uno studio sull'uso delle tecnologie dell'informazione presso i tribunali europei e adottando delle Linee direttrici sulla "cybergiustizia". Nel 2018, vista anche l'evoluzione di questi sistemi e il loro sempre più diffuso utilizzo in ambito giudiziario, ha poi emanato la Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi[146]. Si tratta di un documento programmatico, non vincolante, che enuncia principi applicabili non solo alle autorità pubbliche, ma anche agli attori privati[147].

Lo scopo della Carta è fornire un quadro etico che orienti la "cybergiustizia" verso un suo sviluppo positivo, che effettivamente apporti miglioramenti al settore giudiziario e che possa essere un punto di riferimento per magistrati, autorità pubbliche e utenti. In questo senso, la Carta può essere anche un ottimo strumento di comparazione per valutare le declinazioni e gli esiti a cui la giustizia predittiva approda nei diversi Paesi membri, a seconda delle tecniche impiegate

---

[144] D. Zingales, Risk assessment*: una nuova sfida per la giustizia penale?*, "Diritto penale e uomo – Criminal Law and Human Condition", fascicolo n. 12/2021, p. 2

[145] Così Claudio Castelli, Presidente della Corte d'Appello di Brescia, in un'intervista fatta da Claudia Morelli, in *Altalex.com*, 2017.

[146] Adottata dalla CEPEJ nel corso della sua trentunesima Riunione plenaria, Strasburgo 3 e 4 Dicembre 2018.

[147]Barbaro, op. cit., p. 189.

nei tribunali e nel sistema giudiziario nel suo complesso. La CEPEJ chiarisce anche che è a disposizione degli Stati, dei tribunali e degli attori del diritto in genere per svolgere attività di supporto nell'applicazione dei principi enunciati. La Carta ne sancisce cinque: il principio del rispetto dei diritti fondamentali dell'uomo, il principio di non discriminazione, il principio di qualità e sicurezza, il principio di trasparenza, imparzialità ed equità, ed infine il principio del controllo da parte dell'utilizzatore.

Il rispetto dei diritti fondamentali consiste, nell'ambito della giustizia predittiva, nella necessità che il trattamento di decisioni e dati giudiziari abbia finalità chiare, nel rispetto dei diritti fondamentali garantiti dalla Convenzione europea sui diritti dell'uomo (CEDU) e dalla Convenzione sulla protezione delle persone rispetto al trattamento automatizzato di dati di carattere personale[148]. In particolare, nell'utilizzo dell'IA per dirimere una controversia, per fornire supporto nel processo decisionale giudiziario, o per orientare il pubblico, devono essere garantiti il diritto di accesso a un giudice e il diritto a un equo processo, e dunque il rispetto del principio della parità delle armi e del contraddittorio[149]. Interessante la specificazione sul tipo di approccio da adottare per garantire il rispetto dei diritti fondamentali: esso dovrebbe essere "*etico-fin dall'elaborazione*". Con quest'espressione, viene chiarito nella Carta, si intende un approccio in cui la scelta etica non è demandata all'utilizzatore, ma deve essere stabilita a monte dagli elaboratori del programma. Appare quindi necessaria la predisposizione di un programma che preveda la sinergia tra operatori del diritto, ricercatori in scienze sociali e informatici, attraverso la creazione di équipe pluridisciplinari[150].

Il principio di non discriminazione è il secondo dei cinque individuati dalla CEPEJ: esso consiste nel "prevenire specificamente lo sviluppo o l'intensificazione di qualsiasi discriminazione tra persone o gruppi di

---

[148] Ci si riferisce alla Convenzione n. 108 del Consiglio d'Europa del 1981, come modificata dal Protocollo di emendamento STCE n. 223.

[149] *Carta etica sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi*, primo principio.

[150] Barbaro, op. cit., p. 195.

persone"[151]. Si è già trattato del rischio che la macchina cristallizzi o aggravi discriminazioni che già esistono nella società: per scongiurare questo rischio, occorre operare una vigilanza sia nella fase dell'elaborazione che in quella di utilizzo del sistema, soprattutto quando la metodologia di trattamento dei dati si basa su informazioni direttamente o indirettamente "sensibili", e adottare opportune misure correttive.

Il terzo principio è quello di qualità e sicurezza: è qui che viene espressamente prevista la costituzione di squadre di progetto miste tra i creatori di modelli, i professionisti del diritto e gli studiosi nell'ambito delle scienze sociali, per garantire l'utilizzo, già nella fase di elaborazione del sistema di giustizia predittiva, di "fonti certificate e dati intangibili"[152]. Il significato di questa espressione è da ricercare in una certa modalità di elaborazione dei dati, che deve avvenire sulla base di originali certificati, la cui integrità deve essere garantita in tutte le fasi del procedimento[153]. Nello specifico, ciò significa che le decisioni giudiziarie inserite nel sistema devono avere fonte certa e attendibile, e non possono essere modificate fino a quando non vengano effettivamente utilizzate dal meccanismo di apprendimento. L'elaborazione del modello deve avvenire inoltre in un ambiente sicuro, che garantisca l'integrità e l'intangibilità del sistema.

Il quarto principio, - di trasparenza, imparzialità ed equità, - consiste, oltre che nel "rendere le metodologie di trattamento dei dati accessibili e comprensibili," nella individuazione di autorità o esperti indipendenti che effettuino una verifica sulle metodologie di trattamento, con la concessione di certificazioni o consulenze *ex ante*. Occorre che sia garantito il bilanciamento tra la proprietà intellettuale delle imprese private sui *software*, e l'esigenza di trasparenza - intesa come accesso al processo di creazione del modello -, di imparzialità, di equità e di rispetto degli interessi della giustizia[154]. Una soluzione interessante è

---

[151] *Carta etica sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi*, secondo principio.

[152] *Carta etica sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi*, terzo principio.

[153] Barbaro, op. cit., p. 195.

[154] *Carta etica sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi*, quarto principio.

quella prospettata dallo studio della MSI-NET del Consiglio d'Europa[155], che avanza la possibilità che vengano divulgati al pubblico almeno gli aspetti fondamentali delle informazioni relative agli algoritmi, come le variabili utilizzate dagli stessi, o le quantità e il tipo di dati di apprendimento.

L'ultimo principio enunciato dalla Carta è quello del "controllo da parte dell'utilizzatore". È necessario che l'approccio del modello non sia prescrittivo e che gli utilizzatori siano attori informati, non i destinatari di decisioni la cui motivazione è intellegibile. Per garantire ciò, il linguaggio adottato dal sistema deve rendere chiaro il carattere vincolante o meno delle soluzioni proposte, delle possibilità disponibili, del diritto dell'utilizzatore all'assistenza legale e all'accesso a un tribunale. Inoltre l'utilizzatore ha diritto di essere informato su qualsiasi altro caso precedente al proprio che sia stato trattato dal sistema, e ha diritto di opporsi al trattamento affinché il suo caso venga giudicato direttamente da un tribunale, in applicazione dell'articolo 6 della CEDU[156]. Anche ai professionisti della giustizia vengono accordate alcune garanzie: in particolare dovrebbero avere il diritto di rivedere le decisioni giudiziarie e i dati utilizzati per produrre un risultato, e di potersi discostare dall'esito proposto dall'algoritmo ogni qualvolta lo ritengano opportuno sulla base delle caratteristiche del caso concreto[157].

La Carta, dunque, sottolinea che gli utenti non debbano subire passivamente gli esiti elaborati dai sistemi di giustizia predittiva, e per garantire ciò devono essere adottate tutte le misure necessarie affinché ne venga valorizzata l'autonomia, innanzitutto attraverso l'accesso alle informazioni. In tal senso vi è una differenza tra i Paesi membri dell'Unione Europea e gli Stati Uniti sul diritto di accesso alle modalità di funzionamento degli algoritmi: le autorità giudiziarie statunitensi individuano il punto di bilanciamento tra diritto all'informazione degli utilizzatori e proprietà intellettuale in modo che esso tenda a favorire

---

[155] Si tratta dello studio "*Algorithms and Human Rights – Study on the human rights dimensions of automated data processing techniques and possible regulatory implications*" condotto dal Comitato di esperti sugli intermediari di Internet, un comitato di lavoro del Consiglio d'Europa attivo tra il 2016 e il 2017. In particolare, si fa riferimento alla pagina 38 dello studio.

[156] Si tratta del diritto a un equo processo, sancito dalla Convenzione Europea dei Diritti dell'Uomo.

[157] *Carta etica sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi*, quinto principio.

l'interesse dei privati, mentre il GDPR e *l'AI Act* accordano una più salda tutela a garanzia degli utilizzatori. Il GDPR prevede il diritto dell'interessato a ottenere informazioni sull'esistenza di un processo decisionale automatizzato, compresa la profilazione, e informazioni significative sulla logica utilizzata[158], mentre *l'AI Act* sancisce degli "obblighi di trasparenza per i fornitori e i *deployer* di determinati sistemi di IA"[159], di cui si tratterà più estesamente in seguito.

I sistemi di IA, inoltre, possono essere utilizzati per svolgere attività rivolte allo studio e all'applicazione di metodi statistici volti a predire chi potrebbe commettere un reato, e dove e quando esso potrebbe essere commesso. L'obiettivo di questi sistemi, definiti di polizia predittiva, è quello di impedire la commissione di reati, intervenendo preventivamente. I dati di addestramento inseriti nella macchina possono riguardare qualsiasi elemento attinente alla tipologia, alle modalità e ad ogni altra condizione in cui viene commessa l'azione criminale: possono essere inserite le notizie di reati commessi in precedenza, le informazioni sugli spostamenti e sulle attività di soggetti sospettati, l'analisi delle caratteristiche dei luoghi in cui vengono commessi certi reati, l'individuazione del periodo dell'anno e delle condizioni atmosferiche che sussistono con più frequenza durante la commissione di certi reati. Possono essere inseriti anche, e questo non può che suscitare più d'una perplessità, dati relativi all'origine etnica, al livello di scolarizzazione, alle condizioni economiche, alle caratteristiche somatiche riconducibili a soggetti che con più probabilità potrebbero compiere alcuni tipi di reati[160].

Si possono distinguere due tipi di *software* di polizia predittiva[161]: quelli che individuano le zone, chiamate *hotspots*, in cui più probabilmente verranno commessi alcuni tipi di reati, e quelli che prevedono dove e quando alcuni soggetti potrebbero compiere il prossimo reato, operando quello che viene definito *crime linking*.

---

[158] Articolo 15, 1. (h) del Regolamento UE 2016/679.
[159] Capo IV, articolo 50 del Regolamento UE 2024/1689.
[160] F. Basile, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, "Diritto penale e uomo – Criminal Law and Human Condition", fascicolo n. 10/2019, Diritto penale contemporaneo 2019, p. 10.
[161] Ibid.

Un esempio del primo sistema è il *Risk Terrain Modeling* (RTM): un algoritmo in grado di individuare le aree urbane che con più probabilità saranno teatro di attività criminali, e in particolare di spaccio di sostanze stupefacenti. Viene inserita nel sistema un'enorme massa di dati su fattori ambientali e spaziali associati alla criminalità, come la presenza di luminarie stradali scarse, la vicinanza a locali notturni, a scuole o a stazioni ferroviarie; l'algoritmo opera una mappatura delle aree metropolitane in cui il rischio di spaccio di sostanze stupefacenti è più elevato, con effetti positivi sia sulla programmazione che sull'attuazione di interventi di prevenzione del reato.

L'individuazione degli *hotspots* può però riguardare anche la predizione sul compimento di reati di vario genere, non solo di spaccio: un esempio è *PredPol*, che analizzando il tipo di reato, la data, l'ora e il luogo dello stesso, ha contribuito alla riduzione del tasso di criminalità in diversi Stati nordamericani in cui è stato impiegato. Un esempio italiano è il sistema informatico *X-LAW*, introdotto dalla Questura di Napoli, che elabora un'enorme quantità di dati ricavabili dalle denunce inoltrate alla Polizia di Stato, per individuare le zone e gli orari più ad alto rischio per la commissione di reati.

Per quanto riguarda i sistemi di *crime linking*, invece, essi sono in grado di profilare il possibile autore della serie criminale, - nel caso in cui questo non sia ancora stato individuato, - basandosi sull'idea di fondo che alcuni reati, come le rapine, si compiono in un arco temporale e in una zona geografica molto circoscritti. L'obiettivo di questi *software* è prevedere le azioni che il soggetto attuerà dopo il compimento del reato. Può accadere anche che il soggetto sia già individuato, e il sistema venga impiegato per attività di indagine relative ad azioni criminali che potrebbe aver compiuto prima del reato in occasione del quale egli è stato individuato. Sono presenti esempi di *software* che si ispirano all'idea di *crime linking* in Germania, in Inghilterra, e negli Stati Uniti; in Italia, la Questura di Milano ha elaborato *Keycrime*, che è poi diventato di proprietà di un'azienda privata, di cui si tratterà in seguito.

Accade frequentemente che questi *software* siano di proprietà di aziende private: ciò porta con sé il rischio di opacità circa il funzionamento dell'algoritmo, il quale non viene reso noto dalle aziende. Trattandosi di sistemi che agiscono

nell'ambito della giustizia, e coinvolgono dunque diritti fondamentali, l'assenza di trasparenza e la conseguente impossibilità che vengano eseguiti controlli indipendenti sulla qualità e affidabilità dei risultati prodotti è uno degli aspetti più problematici del loro utilizzo.

Perplessità circa i *software* di polizia predittiva riguardano anche l'assenza di normative organiche che disciplinino le condizioni e i modi di impiego degli stessi, le quali sono unicamente demandate alla prassi degli operatori di polizia, con evidenti rischi circa il rispetto della privacy e del divieto di non discriminazione, affidato unicamente alla sensibilità individuale.

Un loro utilizzo non assistito da idonee garanzie, prima fra tutti la riserva di legge su attività degli organi di polizia che coinvolgono diritti fondamentali, sembra condurre a un modello di "militarizzazione" della sorveglianza[162]. Ancor di più, se si guarda al funzionamento di questi algoritmi, si può facilmente osservare che essi si auto-alimentano, generando un circolo vizioso che porta a un incremento dei controlli in zone già individuate come "calde", a cui seguirà necessariamente una crescita del tasso dei reati in quelle zone a seguito dei controlli più stringenti effettuati dalla polizia; ciò comporta che l'algoritmo valuti quella zona ancora più "calda", con il rischio che altre zone non vengano presidiate dalla polizia, diventando aree franche per la commissione di reati[163].

3.2. Sistemi di *risk assessment*: COMPAS, *Public Safety Assessment* e PATTERN negli Stati Uniti

Nel paragrafo precedente si è tentato di individuare in via generale due delle potenzialità applicative degli algoritmi nell'ambito della giustizia, cioè i casi in cui essi sono impiegati per predire l'esito di una controversia, e i casi in cui il loro utilizzo è volto a impedire la commissione di reati attraverso l'individuazione dei soggetti ritenuti "a rischio" o delle "zone calde".

Tuttavia l'impiego dei sistemi di IA in ambito penale non si riduce alle sole ipotesi di giustizia e polizia predittive sopra menzionate: un campo

[162] Basile, op. cit., p. 13.
[163] Ibid.

particolarmente rilevante, e su cui la sperimentazione è stata senza dubbio florida, è quello del *risk assessment*. Con questa espressione si indicano quei sistemi che operano al fine di ottenere una valutazione prognostica del rischio-recidiva di un imputato e della sua pericolosità sociale[164].

Il loro impiego è stato particolarmente frequente negli Stati Uniti: innanzitutto per la previsione delle recidive, in particolare in materia di libertà vigilata (detta *probation*) e libertà condizionale (*parole*)[165].

Attualmente, risultano in uso circa quattrocento *risk assessment tools*, ma solo quelli di quinta generazione sono da ritenere, secondo la dottrina statunitense, sistemi di IA. Questi ultimi infatti funzionano sulla base di processi di *machine learning*; sono cioè dotati della capacità di autoapprendimento. Essi possono evolversi indipendentemente dal controllo umano, e sulle loro procedure e modalità di funzionamento vige ancora segretezza; non è dunque possibile escludere che questi sistemi siano già utilizzati quali *criminal risk assessment tools*. È fuor di dubbio invece l'impiego in ambito penale dei *Risk and Needs Assessment tools*, cioè i sistemi di quarta generazione: essi si caratterizzano per un dataset ampio che garantisce una maggiore predittività, e per la predisposizione di programmi volti a diminuire il rischio di recidiva e il reinserimento del soggetto[166].

Valutare quale sia il grado di probabilità con cui un soggetto, già artefice di un reato, possa commetterne un altro, è sempre stata un'esigenza del diritto penale sostanziale e processuale. Il rischio di recidiva è infatti uno dei temi centrali che più urgentemente si pongono all'attenzione dei giudici, per le ovvie ricadute che hanno sull'applicazione di misure di sicurezza, misure cautelari e misure di prevenzione, o ancora sulla concessione della sospensione condizionale di una pena o l'affidamento in prova al servizio sociale[167]. La decisione sulla pericolosità sociale di un soggetto è spesso rimessa alla valutazione di periti, e

---

[164] Zingales, op. cit., p. 2.
[165] Falletti, op. cit., p. 216.
[166] Zingales, op. cit., p. 4.
[167] F. Basile, *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, in F. Basile, M. Caterini, S. Romano (a cura di), Il sistema penale ai confini delle *hard sciences*: Percorsi epistemologici tra neuroscienze e intelligenza artificiale, Pacini Giuridica 2021, p. 18.

in ogni caso alla discrezionalità dei giudici, che valutano secondo buon senso; per questo in molti sostengono che l'impiego di *risk assessment tools* possa garantire l'elaborazione di "valutazioni attuariali," attraverso la rielaborazione di "quantità enormi di dati al fine di far emergere relazioni, coincidenze, correlazioni, che consentano di profilare una persona e prevederne i successivi comportamenti, anche di rilevanza penale[168]".

Si distinguono, sulla base della finalità che perseguono, due tipi di *risk assessment tools*. I primi, detti *pretrial risk assessment instruments*, sono volti alla valutazione dell'eventuale sussistenza delle condizioni per il mantenimento della carcerazione preventiva; i secondi, detti *risk assessment instruments*, valutano il rischio di recidiva o l'ammissibilità di misure alternative alla detenzione[169].

L'impiego di questi strumenti ha incontrato, a ragione, più di una reticenza nel nostro Paese: affidare la valutazione circa l'applicazione di misure che limitano la libertà personale ad un algoritmo appare questionabile sotto molteplici aspetti. Innanzitutto il suo funzionamento, si è detto, non sempre è intellegibile; in secondo luogo tali sistemi possono cristallizzare e perpetrare discriminazioni che già sono presenti nella società, attraverso dei *bias* spesso non individuabili, proprio per via dell'assenza di trasparenza.

Sebbene anche negli Stati Uniti parte della dottrina abbia sollevato perplessità circa la compatibilità del loro impiego con i principi del *fair trial*, sanciti nel V e nel XIV emendamento alla Costituzione nordamericana, la giustizia nordamericana fa ampio utilizzo di sistemi di *risk assessment*. Occorre precisare che la valutazione su *se* e *come* impiegare questi strumenti è demandata all'organo giudicante: il loro impiego è raccomandato, non obbligatorio, e dunque ben lontano dall'essere sostitutivo del ruolo del giudice[170]. In ogni caso, si può affermare che negli ultimi anni la loro diffusione è stata inarrestabile: a riprova di ciò si noti la produzione normativa in materia, che tra il 2012 e il 2015

[168] Ibid.
[169] Zingales, op. cit., p. 5.
[170] Zingales, op. cit., p. 6.

è stata particolarmente intensa, con l'emanazione di venti leggi in ben quattordici Stati[171].

Gli algoritmi predittivi impiegati nelle giurisdizioni nordamericane sono molto eterogenei, e tengono in considerazione fattori di rischio sia "statici" che "dinamici". I fattori appartenenti alla prima tipologia sono quelli che permangono per tutto il corso della vita del soggetto, come il genere e l'età del primo arresto; mentre i fattori della seconda categoria possono variare, - o non sussistere più, - con il trascorrere del tempo, come l'età, il lavoro, l'utilizzo di sostanze psicotrope. Possiamo ulteriormente distinguere gli algoritmi in "statali", cioè elaborati dai governi o comunque con la partecipazione degli stessi, e "privati", su cui vige il segreto industriale.

Una vicenda d'oltreoceano che ha assunto un ruolo centrale nel dibattito giurisprudenziale e dottrinale è sicuramente il caso *Loomis v. Wisconsin*[172]. Nel Febbraio del 2013, Eric Loomis è stato arrestato mentre guidava un'auto usata in una sparatoria. Poco dopo il suo arresto, si era dichiarato colpevole in relazione a due dei cinque capi d'accusa: guida del veicolo senza il consenso del proprietario e oltraggio a un pubblico ufficiale[173]. La corte locale lo aveva condannato a sei anni di reclusione e a cinque anni di *extended supervision,* negando il riconoscimento della libertà condizionale, anche sulla base del giudizio di alto rischio fornito da un sistema di valutazione del rischio di quarta generazione. In particolare, il *software* impiegato per la valutazione fu COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*), sviluppato dalla società privata *Northpointe, Inc.* (ora *Equivant*), la quale, occorre sottolinearlo, anche successivamente alla vicenda ha rifiutato di rendere note le modalità di funzionamento del sistema, adducendo la sussistenza di segreto industriale[174].

---

[171] M. Gialuz, *Quando la giustizia penale incontra l'Intelligenza Artificiale: luci e ombre dei* risk assessment tools *tra Stati Uniti ed Europa*, "Diritto Penale Contemporaneo" 2019, p. 4.
[172] Si tratta della decisione *State of Wisconsin v. Eric L. Loomis*, 881 N.W.2d 749 (Wis. 13 Luglio 2016).
[173] Falletti, op. cit., p. 219.
[174] R. Perona, *Giustizia artificiale intelligente? Spunti per il costituzionalista europeo a partire dall'esperienza colombiana*, in federalismi.it, n. 26/2025, p. 109.

L'imputato adiva, a seguito del rigetto di un'istanza di *post-convinction release*[175], la Corte Suprema del Wisconsin: in primo luogo, il difensore lamentava la violazione del diritto del suo assistito ad essere valutato sulla base di informazioni accurate, in secondo luogo la violazione del diritto ad una sentenza individualizzata, e infine l'appartenenza al genere maschile tra i vari dati utilizzati per valutare la pericolosità[176].

La Corte era dunque chiamata a decidere "se l'uso della valutazione del rischio mediante sistema COMPAS violi il diritto al giusto processo di un imputato, sia perché la natura proprietaria di COMPAS impedisce di contestare la validità scientifica della valutazione del sistema, sia perché le valutazioni di COMPAS tengono conto del genere[177]".

Il giudice del Wisconsin ha rigettato i dubbi di costituzionalità relativi alla violazione del *due process* dell'imputato: si è espresso nel senso di negare che l'utilizzo di COMPAS avesse comportato, nel caso di specie, una violazione del diritto a un giusto processo, anche se le modalità di funzionamento del *software* non erano note né all'imputato né al giudice. In particolare, ha sostenuto che l'utilizzo di COMPAS nella fase di condanna fosse "utile per fornire alla corte, in sede decisoria, quante più informazioni possibili al fine di giungere a una sentenza personalizzata"[178]. È stata dunque confermata la condanna di Loomis a sei anni di carcere, chiarendo l'ammissibilità di una decisione che sia in parte fondata sulla valutazione algoritmica del rischio di recidiva[179].

La giurisprudenza nordamericana giustifica il mantenimento dell'opacità in virtù della tutela dell'*intellectual property*, che nel caso di specie agisce come protezione degli interessi economici di chi sviluppa la tecnologia di *machine learning*: tale orientamento non appare condivisibile, perché costituisce un'indebita compressione del diritto di difesa, che deve essere sempre garantito

---

[175] Si tratta della richiesta di una nuova udienza per la revisione della condanna.
[176] Gialuz, op. cit., p. 6.
[177] *State of Wisconsin v. Eric L. Loomis*, 881 N.W.2d 749 (Wis. 13 Luglio 2016), par. 6, traduzione di R. Perona, op. ct., p. 109.
[178] *State v. Loomis*, cit., parr. 8 e 72, traduzione di R. Perona, op. cit., p. 109.
[179] Perona, op. cit., p. 107.

in una prospettiva di bilanciamento con il dispiegamento delle potenzialità delle tecnologie di IA e di tutela del segreto industriale[180].

Il sistema COMPAS, e la pronuncia che ha affermato l'opportunità del suo impiego per la valutazione del rischio di recidiva, sono stati ampiamente criticati. In primo luogo, l'organizzazione ProPublica ha sostenuto che il *software* contenesse dei *bias* discriminatori nei confronti delle persone nere, poiché alcuni fattori dinamici oggetto di valutazione erano strettamente correlati alla razza. È stato accertato che le persone nere venivano valutate dal sistema come soggetti che più probabilmente avrebbero potuto compiere reati, in misura pari al doppio rispetto alle persone non nere[181].

In secondo luogo, pare difficile comprendere perché la circostanza che l'algoritmo non fosse conosciuto non sia stata ritenuta sufficiente ad escluderne l'affidabilità. Il fatto che l'algoritmo fosse brevettato e segreto implica che la ratio alla base del risultato prodotto dal sistema, - che ha concorso, seppur insieme ad altri elementi, alla decisione di condanna, - sia sconosciuta, impossibile da sottoporre al vaglio critico della riflessione del giudice. Non si può quindi escludere che COMPAS abbia attribuito rilevanza a fattori apparentemente neutri, ma in realtà strettamente legati all'appartenenza a una certa categoria, compiendo in questo senso una discriminazione razziale.

Per contro, la Corte Suprema del Wisconsin ha sostenuto che l'affidabilità del sistema potesse essere verificata dall'imputato attraverso il manuale d'uso dello strumento: confrontando, in particolare, i dati individuali (cioè gli *input*) e le valutazioni di rischio finali (*output*)[182]. Successivamente, Loomis ha presentato domanda di revisione giudiziaria, il *writ of certiorari*, alla Corte Suprema degli Stati Uniti, la quale è stata respinta[183].

Benché non condivisibile, la decisione della Corte contiene alcune precisazioni importanti circa il rapporto uomo-macchina nell'ambito della giustizia predittiva. Le riflessioni elaborate sono riconducibili, genericamente, al principio dello *human in the loop*, poiché vengono chiariti i criteri a cui dovrebbe

---

[180] Falletti, op. cit., p. 221.
[181] Gialuz, op. cit., p. 5.
[182] Gialuz, op. cit., p. 6.
[183] Certiorari denied, *Loomis v. Wisconsin*, 137 S. Ct. 2290 (2017).

sottostare il giudice che utilizzi sistemi di IA nell'esercizio della funzione giudicante. Innanzitutto, la Corte individua tre ipotesi in cui l'impiego dell'algoritmo COMPAS è giustificato: la verifica della possibilità di indirizzare un condannato "a basso rischio" verso una pena non detentiva; la valutazione sulla necessità che un condannato debba essere sottoposto a misure di sorveglianza; la definizione dei termini e delle condizioni della libertà vigilata. Tuttavia, la sussistenza di una di queste ipotesi non è sufficiente, secondo la Corte, a rendere legittimo l'impiego di COMPAS: vi è in capo al giudice un obbligo di motivazione, in modo da assicurare che la decisione non sia fondata esclusivamente sul risultato prodotto dal *software*, ma che quest'ultimo sia solo uno degli elementi che, insieme ad altri, concorre alla formazione del suo convincimento[184].

Questa pronuncia ha contribuito a consolidare la posizione di parte della dottrina secondo cui gli algoritmi predittivi possono determinare il perfezionamento delle decisioni del giudice[185], integrando la sua esperienza, che in quanto umana è limitata, con l'operato della macchina, che può invece attingere da un enorme bacino di dati. Questa letteratura si muove anche dal presupposto che tali strumenti potrebbero contribuire a contrastare il fenomeno della "*mass incarceration*", molto sentito negli Stati Uniti, che sono il Paese con il maggior numero di detenuti su scala mondiale[186].

In ogni caso, sarebbe erroneo affermare che l'incertezza sull'effettiva imparzialità degli algoritmi predittivi sia stata priva di conseguenze nel dibattito statunitense: interessante in questo senso l'esito che ha avuto in California la *Proposition 25* sul *replacement* del *cash bail* con i *pretrial risk assessment tools*. In particolare il quesito posto ai californiani riguardava la sostituzione della legge sul *cash bail*, la quale prevede il rilascio dell'imputato dietro pagamento di una cauzione in attesa del processo, con l'entrata in vigore del *California Money Bail Reform Act*. Esso prevedeva il mantenimento della carcerazione preventiva nel caso in cui un algoritmo ne avesse indicato l'opportunità

[184] Nardocci, op. cit., p. 429.
[185] Gialuz, op. cit., p. 7.
[186] Zingales, op. cit., p. 6.

attraverso la classificazione in "alto", "medio" e "basso" del rischio di recidiva[187]. La popolazione si è espressa in senso sfavorevole all'abrogazione del *cash bail*: si tratta di una scelta significativa, che esprime una forte diffidenza verso i *risk assessment tools* quando ad essere coinvolta è la libertà personale, anche a discapito delle prospettive di maggiore precisione ed obiettività che questi algoritmi dovrebbero garantire. Ancor di più se si considera che ad essere favorita è la legge sul *cash bail*, già portatrice di non poche criticità in termini di rispetto del principio di eguaglianza sostanziale: è evidente infatti la discriminazione che essa produce nei confronti dei soggetti meno abbienti, che non possono sostenere economicamente la cauzione, e a cui quindi è precluso il beneficio del rilascio in attesa della celebrazione del processo.

Il caso COMPAS ha costituito un precedente che ha messo in evidenza le criticità dell'impiego degli algoritmi predittivi per valutare il rischio di recidiva degli imputati. Successivamente, sono stati creati altri *tools* con l'obiettivo di superare le problematiche attinenti all'assenza di trasparenza e alla produzione di effetti discriminatori del *software*. In particolare, si è affermata la necessità che dovessero essere esclusi dai dati inseriti nel sistema quelli attinenti alle condizioni economiche, razziali e di genere.

Nel 2018 viene rilasciato pubblicamente, negli Stati Uniti, il *Public Safety Assessment* (PSA), sviluppato dalla *Laura and John Arnold Foundation.* Per il suo addestramento sono stati utilizzati i *reports* di settecentocinquantamila casi, attinenti oltre trecento giurisdizioni nordamericane. Il funzionamento dell'algoritmo è pubblico: esamina nove fattori, relativi all'età, all'imputazione e ai precedenti penali del soggetto. La finalità del PSA consiste nell'individuare due fattori di rischio: il pericolo che il soggetto non si presenti in udienza, e la probabilità che commetta un reato se rilasciato prima del dibattimento. I dati impiegati sono neutrali, e non tutti i parametri hanno lo stesso peso; tuttavia, il rischio della presenza di *bias* razziali non pare del tutto scongiurato[188].

Uno dei *software* più recenti è stato elaborato dal *Department of Justice* (DOJ) nel 2019: si tratta di un sistema di giustizia predittiva di natura istituzionale, che

---

[187] Zingales, op. cit., pp. 7-8.
[188] Gialuz, op. cit., p. 7.

nella fase creativa ha visto il coinvolgimento di portatori di interessi privati e pubblici. Il sistema è denominato PATTERN (*Prisoner Assessment Tool Targeting Estimated Risk and Needs*), e ha incontrato il favore, - quantomeno parziale, - della dottrina penale statunitense, perché garantisce un certo grado di trasparenza, che i precedenti algoritmi di giustizia predittiva non sono stati in grado di garantire.

Tuttavia, anche PATTERN è ben lontano dal rispetto di standard adeguati ad un *software* finalizzato a compiere valutazioni che coinvolgono diritti fondamentali dell'individuo: in particolare, è stato criticato l'utilizzo da parte di PATTERN di dati elaborati da un altro sistema, denominato BRAVO (*Bureau Risk and Verification Observation*), il cui funzionamento non è stato reso pubblico. Non è quindi possibile sapere se PATTERN sfrutti le potenzialità del *machine learning* per il calcolo del rischio di recidiva, dato che dei diciassette *items* che prende in considerazione alcuni provengono dal poco conosciuto BRAVO. Sebbene dunque sia classificato tra i sistemi più evoluti di quarta generazione, non si può escludere che il supporto di BRAVO ne determini la qualificazione come algoritmo di autoapprendimento, e quindi di quinta generazione.

Inoltre, ci sono dubbi anche circa l'effettiva imparzialità del sistema: secondo uno studio preliminare svolto dal *Brennan Center of Justice*, oltre la metà di tutti gli uomini neri sono stati valutati dal sistema come ad alto rischio di recidiva; ciò suggerisce che gli *outputs* generati da PATTERN siano inficiati da giudizi discriminatori. Questo è ancor più rilevante se si pensa che, a differenza di quanto avviene con COMPAS ed altri *risk assessment tools*, il cui giudizio è solo uno degli elementi che concorrono a formare il convincimento del giudice, il risultato prodotto da PATTERN è l'unico su cui si basa la decisione.

Il carattere decisivo del livello di rischio di recidiva attribuito dall'algoritmo è stato stabilito dal *First Step Act*[189], che ha anche previsto l'istituzione di una *Independent Review Committee*, atta a compiere delle verifiche annuali sull'affidabilità del sistema.

[189] Public Law No. 115-391 del 21/12/2018.

Pur essendoci più di una remora in tema di trasparenza del sistema, PATTERN consiste comunque in un passo in avanti rispetto ai sistemi precedenti. Esso è programmato per prevedere il rischio di recidiva entro tre anni dalla liberazione, e opera una distinzione tra "*violent recidivism*" e "*general recidivism*", oltre a distinguere tra il rischio di recidiva delle donne e quello degli uomini. Soprattutto, PATTERN è in grado di garantire una maggiore attendibilità dei risultati, confrontati con altri programmi di *risk assessment*: in particolare, si stima che i risultati prodotti siano più accurati del 15% rispetto a quelli prodotti da altri *tools*[190].

3.3. I rischi e i benefici dei sistemi di giustizia predittiva: *Keycrime* come esempio virtuoso in Italia

Alla luce di quanto detto circa le caratteristiche dei sistemi di giustizia predittiva, appare doverosa una riflessione circa la compatibilità del loro impiego con il rispetto delle garanzie costituzionali – quali la riserva di legge, il principio del giudice naturale predefinito per legge, imparziale e terzo, il diritto a un equo processo, il diritto di difesa, il diritto alla motivazione della decisione.

Innanzitutto, è fuor di dubbio che il ruolo attribuito all'algoritmo nel processo decisionale del giudice che miri a stabilire il livello di pericolosità sociale debba essere stabilito dalla legge, che ne individua i casi e i modi di impiego. In particolare, la normativa europea esclude che la decisione possa discendere esclusivamente da un processo automatizzato, senza che esso sia sottoposto alla supervisione umana[191].

Le criticità emergono innanzitutto con riferimento ai caratteri di opacità e di segretezza che accomunano tutti gli algoritmi proprietari, poiché la tutela della proprietà intellettuale si pone in tensione, come si è visto nel caso COMPAS, con il principio di trasparenza.

Pare più complesso, invece, fugare ogni dubbio circa l'assenza di *bias* nel funzionamento del sistema adottato. Si è visto come COMPAS per primo, ma anche PATTERN, siano caratterizzati da una certa opacità di funzionamento.

---

[190] Zingales, op. cit., pp. 9-13.

[191] C. Parodi, V. Sellaroli, *Sistema penale e intelligenza artificiale: molte speranze e qualche equivoco*, in "Diritto Penale Contemporaneo", fascicolo n. 6/2019, p. 67.

Non è dunque possibile escludere che i risultati generati da questi sistemi siano completamente scevri da giudizi discriminatori avverso categorie deboli. L'effettiva democraticità del sistema dipende in larga misura dai dati che vengono inseriti quali *input*, che costituiscono la base dello strumento. Occorre tenere in considerazione che i dati sono "una risorsa inesauribile, riutilizzabile, facilmente trasportabile e aggredibile, duplicabile, durevole e condivisibile[192]", e dunque il loro valore può variare notevolmente a seconda delle finalità per cui vengono utilizzati: la qualità dei dati è fondamentale nella fase creativa del modello, poiché è nel momento dell'addestramento che si determina l'imparzialità del sistema. Tanto più i dati sono "di buona qualità, e rispondenti alla realtà da interpretare e prevedere[193]", tanto maggiore sarà l'efficacia predittiva dell'algoritmo. Tuttavia spesso i programmatori non dispongono di una quantità sufficiente di dati "diretti", cioè immediatamente pertinenti ai comportamenti rilevanti ai fini dell'addestramento, e dunque si servono di *proxy data* e dati "indiretti" in genere.

L'esito di questa operazione è l'immissione nel sistema di dati relativi, ad esempio, all'etnia, al genere, alla zona di residenza e al grado di scolarizzazione: il rischio è che il modello predittivo individui correlazioni rilevanti tra queste informazioni, elaborando un giudizio negativo, assegnando così un livello "alto" di rischio per via della valutazione di elementi che non hanno nulla a che vedere con la probabilità di recidiva.

Se è vero che, nel caso COMPAS, la Corte Suprema del Wisconsin ha precisato che il giudice debba, nel svolgere il suo ruolo decisionale, bilanciare il risultato del programma con altri fattori – e in questo senso nega qualsiasi compressione dell'esercizio della discrezionalità dell'organo giudicante -, occorre anche considerare che con PATTERN il governo nordamericano ha fatto dell'algoritmo l'unico elemento atto a determinare la decisione del giudice, senza per contro accordare un sufficiente grado di trasparenza del sistema.

La presenza di automatismi è inammissibile nell'ordinamento italiano, perché contraria alla tutela del principio del giudice naturale, all'obbligo di

[192] Ibid.
[193] Ibid.

motivazione, al diritto di difesa. Non è inoltre compatibile col principio del giusto processo, come disciplinato all'art. 111 Cost. e all'art. 6 della Convenzione europea per la salvaguardia dei diritti umani e delle libertà fondamentali, oltre che con il disposto dell'art. 2 GDPR, il quale garantisce il diritto alla spiegazione della decisione automatizzata[194].

L'iter decisionale non deve, quindi, essere completamente automatizzato: a tal proposito è stato affermato, per la prima volta dalla Corte costituzionale colombiana, il principio dello *human oversight*. La Corte è stata chiamata a pronunciarsi sull'impiego di *ChatGPT* nelle aule giudiziarie: la vicenda[195] riguardava l'utilizzo del sistema di *machine learning* da parte del giudice, accusato di essersene servito per la definizione integrale di un processo riguardante un minore con disabilità. In quell'occasione il giudice costituzionale ha chiarito che le tecnologie di IA possono costituire solo un ausilio secondario del giudice, il quale deve sempre supervisionare sullo strumento, indipendentemente dal grado di complessità e di indipendenza dello stesso da *input* esterni[196]. Non si tratta dunque di una condanna dell'IA: viene invece sottolineata l'esigenza di predisporre linee guida per il giudice, quando nell'adempimento delle proprie funzioni si serva di algoritmi. La Corte chiarisce che l'esercizio dell'attività giurisdizionale non può prescindere dal rispetto di *good practise* e standard di utilizzo delle tecnologie, da garantire anche attraverso programmi di formazione che devono essere predisposti dall'organo di autogoverno della magistratura colombiana.

Ciò che emerge dall'orientamento della Corte costituzionale colombiana è l'accoglimento di sistemi di IA in ambito giudiziario, ma non come "un nuovo umano, bensì un supporto tecnico sprovvisto di quelle qualità – la ragionevolezza, per citarne una -, che rendono cruciale il rispetto del principio dello *human oversight*"[197].

---

[194] Falletti, op. cit., p. 217.
[195] La pronuncia che è stata poi oggetto della decisione della Corte costituzionale colombiana è la sentenza n. 032 del 30 Gennaio 2023, resa dal *Primo Laboral* del Circuito di Cartagena.
[196] Nardocci, op. cit., p. 447.
[197] Nardocci, op. cit., pp. 448-450.

Il rispetto del diritto a un equo processo, - che è uno dei capisaldi di ogni ordinamento democratico, - è messo a rischio anche quando vi sia un utilizzo di *proxy data*, impiegati per sostituire i dati diretti quando non sono disponibili, o di valutazioni *bucket*. Queste ultime sono valutazioni in cui vengono inseriti dati che non sono riferibili alla persona da valutare, ma alla categoria a cui appartiene[198]: l'effetto distorsivo in questo caso è massimo, poiché il soggetto può addirittura essere valutato sulla base di informazioni non proprie, ma perché classificato dal modello come appartenente ad una certa categoria definita "a rischio". La pertinenza dei dati, la loro qualità, provenienza, non manipolabilità: queste sono le condizioni necessarie per la creazione di uno strumento che sia davvero di ausilio ai giudici, senza un'ingiustificata compressione dei diritti fondamentali dell'imputato.

Un'attenta e regolamentata selezione dei dati inseriti quali *input*, dunque, e una statalizzazione, o quantomeno la garanzia di una partecipazione dello Stato alla creazione dei sistemi di giustizia predittiva, sono i risultati da conseguire per un'attenuazione dei rischi più urgenti legati all'impiego dell'IA nella giustizia penale.

Con quanto esposto fino ad ora si sono volute indicare le perplessità che la c.d. "giustizia algoritmica" ha sollevato per via di alcune caratteristiche insite nei modelli, come l'opacità dei processi logici e la riproduzione di giudizi discriminatori della realtà che è chiesto al sistema di rappresentare. Tuttavia, non mancano gli esempi virtuosi e le opportunità per il loro impiego.

Si può citare in questo senso *Keycrime*, programma di polizia preventiva e di polizia giudiziaria elaborato dalla Questura di Milano, il quale si caratterizza per la sua utilizzabilità solo a fronte di condotte seriali. Il suo limite dunque, che ne determina però anche la compatibilità con il nostro sistema penale, è quello di non operare nelle ipotesi di occasionalità nella commissione del reato, o di fatti criminali potenzialmente seriali, all'atto della loro insorgenza.

Il programma è stato progettato per contrastare le rapine, tuttavia può essere applicato a tutti i reati seriali con modalità operative simili, quali truffe agli

[198] Paroli, Sellaroli, op. cit., p. 69.

anziani, furti in appartamento, violenze sessuali e altri. L'efficacia dei risultati dipende da fattori quali la quantità e qualità dei dati, la tempestività con i quali vengono raccolti, la completezza dell'inserimento delle informazioni. Nel *database* vengono inserite le dichiarazioni delle vittime, raccolte tempestivamente attraverso moduli standardizzati, oltre ad immagini di sorveglianza, tracce biologiche, rilievi, e altri elementi investigativi. Il sistema compie un confronto costante di tutte le informazioni, al fine di individuare i luoghi in cui avvengono i reati, e classificarli in base alle caratteristiche dell'autore. La comparazione consiste nell'individuazione di similitudini e differenze rispetto ai casi precedenti, prendendo in considerazione moltissimi elementi, come l'abbigliamento dei soggetti, le modalità operative, le espressioni linguistiche utilizzate, gli orari in cui si consuma il fatto criminale, le modalità di fuga.

Questo sistema, i cui risultati sono stati soddisfacenti, consente di mappare le serie criminali, e prevedere luogo e orario in cui potrebbero essere commessi reati in futuro. Sulla base delle previsioni formulate da *Keycrime* vengono organizzati servizi di appostamento per cogliere in flagranza il reo o prevenire la commissione del reato: in questi casi, sussistono gli elementi sufficienti per avviare un procedimento penale.

Il pregio del programma è riconducibile al rispetto del principio di non discriminazione: esso è garantito dal fatto che sia assicurata la sorveglianza umana su ogni fase del processo, e dalla formulazione di una previsione basata su molteplici elementi e non solo su dati statistici. Inoltre *Keycrime* non opera attraverso meccanismi di autoapprendimento, i quali, si è visto, sono spesso forieri di *bias* per via delle correlazioni individuate tra fattori che producono elementi discriminatori. La compatibilità del sistema con l'ordinamento dipende dal fatto che i dati siano stati acquisiti nel rispetto delle norme del codice di procedura penale; ogni dato deve essere utilizzabile come prova o elemento investigativo e deve poter essere prodotto davanti al giudice.

Effettuato il controllo sulla legittimità dei dati, occorre valutare i criteri e i metodi con cui l'algoritmo compie le correlazioni tra gli stessi, e quando queste correlazioni possono essere ritenute sufficienti[199].

In ogni caso, la dottrina ha individuato tre specifiche esigenze che i sistemi di *risk assessment* sono in grado di soddisfare: due, più generali, si collocano su un piano di valutazione dei costi e dei benefici e di accesso alle risorse in materia di pubblica sicurezza e di tutela legale, mentre la terza è posta su un piano individuale[200]. Per quanto attiene all'efficienza dell'amministrazione della giustizia, i vantaggi riguardano l'elaborazione in modo automatizzato di una pianificazione strategica di ampio respiro sul contrasto alle attività criminali, con la programmazione delle previsioni e delle priorità di azione, e l'inserimento, nella valutazione dei profili, di specifiche attività relative alla riduzione del crimine. In un'ottica individuale, il vantaggio consiste nella valutazione della previsione di recidiva[201].

In conclusione, l'utilizzo dell'IA a fini predittivi in un contesto giudiziario può apportare benefici all'intero sistema: la presa di consapevolezza circa i rischi dell'impiego dell'IA in ambito penale, lungi dal consistere in una demonizzazione dei sistemi di *risk assessment*, è funzionale al garantirne un impiego che sia illuminato dalle garanzie e dai principi fondamentali dell'ordinamento, al fine di una realizzazione delle loro potenzialità, che apporti incrementi di efficienza.

---

[199] Paroli, Sellaroli, op. cit., pp. 56-59.
[200] Falletti, op. cit., p. 217.
[201] Ibid.

Capitolo quarto
Un'analisi comparatistica dei modelli di disciplina

4.1. Il Regolamento UE 2024/1689 sull'intelligenza artificiale

L'*AI Act* dispone regole armonizzate sull'intelligenza artificiale; è stato emanato nel giugno del 2024 dal Parlamento europeo e dal Consiglio dell'Unione Europea. L'urgenza di una regolamentazione della materia si è resa necessaria, come si può evincere dai Considerando del Regolamento[202], per via dei numerosi settori dell'economia interessati dall'impiego di IA, e per la notevole influenza che tali tecnologie hanno in vario modo esercitato in molti ambiti della società. Si comprende ancor di più la rilevanza della materia se si considera che i sistemi di IA hanno rilievo transfrontaliero e possono facilmente circolare attraverso il territorio dell'Unione Europea.

Al Capo I sono contenute le disposizioni generali, che attengono l'ambito di applicazione della disciplina; ancor prima, l'articolo 1 in particolare, chiarisce lo scopo del regolamento. L'emanazione della normativa risponde all'esigenza di "migliorare il funzionamento del mercato interno e promuovere la diffusione di un'intelligenza artificiale (IA) antropocentrica e affidabile, garantendo nel contempo un livello elevato di protezione della salute, della sicurezza e dei diritti fondamentali sanciti dalla Carta dei diritti fondamentali dell'Unione europea, compresi la democrazia, lo Stato di diritto e la protezione dell'ambiente, contro gli effetti nocivi dei sistemi di IA nell'Unione, e promuovendo l'innovazione"[203]. Da un lato quindi, la tutela dei diritti e delle libertà fondamentali, dall'altra la rimozione degli ostacoli allo sviluppo e alla diffusione di queste tecnologie[204]. Per via del carattere transfrontaliero dell'impiego dei sistemi di IA, la sola regolamentazione nazionale, generatrice di divergenze tra gli Stati membri,

---

[202] In particolare, *Regolamento (UE) 2024/1689*, in Gazzetta Ufficiale dell'Unione Europea, Considerando 3 e 4.
[203] *Regolamento (UE) 2024/1689*, in Gazzetta Ufficiale dell'Unione Europea, art. 1 par. 1.
[204] M. Pappalardo, *L'ambito di applicazione dell'*AI Act, Navigare *l'European AI Act*, AIRIA Associazione per la Regolazione dell'Intelligenza Artificiale (a cura di), Wolters Kluwer, Milano 2024, p. 30.

avrebbe determinato la frammentazione del mercato dell'Unione e rischi per la certezza del diritto, e sarebbe stata dunque del tutto inadeguata.

Il secondo paragrafo dell'art. 1 definisce l'oggetto del regolamento, e in particolare: l'*AI Act* dispone regole armonizzate per l'immissione sul mercato, la messa in servizio e l'uso dei sistemi di IA nell'Unione, oltre a individuare le pratiche vietate, nonché alcuni requisiti richiesti specificatamente per i sistemi definiti "ad alto rischio", e gli obblighi per gli operatori di tali sistemi. Vengono inoltre sancite regole di trasparenza, per l'immissione sul mercato di modelli di IA definiti "per finalità generali", oltre a disposizioni in materia di monitoraggio e vigilanza del mercato, governance europea e nazionale, ed esecuzione. Infine, sono disposte misure incentivanti che favoriscano l'innovazione, con particolare riguardo alle PMI, ivi comprese le start-up[205].

L'ambito oggettivo di applicazione del Regolamento è stabilito all'art. 3, che enumera una serie di definizioni atte a circoscrivere il perimetro dell'*AI Act.* Viene fornita[206] la nozione di IA accolta dal Regolamento: si deve trattare di un sistema automatizzato, con un certo grado di autonomia e dotato di meccanismi di autoapprendimento, che produca un risultato – detto *output*, - che sia generato a partire da un processo di deduzione a partire dall'*input.* Dunque le disposizioni del Regolamento si applicano unicamente ai sistemi che abbiano tutti gli elementi appena indicati, mentre gli altri programmi informatici non sono soggetti ai divieti e agli obblighi disposti dalla normativa.

Interessante, al n. 63 dell'elenco, è la definizione di modelli di IA caratterizzati da *general purpose technology*[207]. Con questa espressione si fa riferimento a quel tipo di tecnologie, come l'elettricità o il computer, che apportano enormi benefici alla società, sia in ambito economico che ambientale, con un impatto significativo sulle attività industriali e sociali, ma che proprio per la loro pervasività e influenza in molti ambiti della vita, possono comportare rischi per interessi di rilevanza pubblica e per i diritti fondamentali[208].

---

[205] *Regolamento (UE) 2024/1689*, in Gazzetta Ufficiale dell'Unione Europea, art. 1 par. 2.
[206] *Regolamento (UE) 2024/1689*, in Gazzetta Ufficiale dell'Unione Europea, art. 3 n. 1.
[207] *Regolamento (UE) 2024/1689*, in Gazzetta Ufficiale dell'Unione Europea, art. 3 n. 63.
[208] Pappalardo, op. cit., p. 29.

Il Regolamento dunque individua dei modelli di IA, definiti "per finalità generali": le previsioni per questo tipo di modelli sono applicabili nel caso in cui essi siano addestrati con una grande quantità di dati, abbiano carattere generale e siano in grado di svolgere tanti compiti distinti, - a prescindere dalle modalità di immissione nel mercato - e, infine, siano suscettibili di essere integrati in una varietà di sistemi o applicazioni a valle. Alcuni di questi modelli di IA per finalità generali sono a "rischio sistemico", e ad essi si applicano delle disposizioni specifiche per via della grande rilevanza che rivestono nell'ambito dell'Unione e dei rischi legati ad un loro impiego non conforme al Regolamento.

Sotto il profilo soggettivo, il Regolamento chiarisce che l'*AI Act* si applica a tutti gli operatori che abbiano un ruolo nelle varie fasi, che si tratti della fase dello sviluppo del sistema, della sua immissione sul mercato unionale, fino all'utilizzazione degli stessi. L'art. 3 definisce anche le figure che svolgono attività legate all'impiego dell'IA, individuate nel Regolamento: il fornitore, il *deployer*, il rappresentante autorizzato, l'importatore e il distributore.

Il fornitore è una persona fisica o giuridica, un'autorità pubblica, un'agenzia o un altro organismo sviluppatore di un sistema di IA o di un modello di IA per finalità generali, e che immette sul mercato tale sistema, o che mette in servizio il sistema di IA con il proprio marchio, a titolo oneroso o gratuito. Il *deployer,* è una persona fisica o giuridica, un'autorità pubblica, un'agenzia o un altro organismo che utilizza un sistema di IA sotto la propria autorità, salvo l'ipotesi in cui il sistema di IA sia utilizzato da persone fisiche nel corso di un'attività personale non professionale. Vale dunque, come nel GDPR, la c.d. *household exemption*, cioè l'esclusione dall'applicazione delle norme sulla protezione dei dati personali per l'esercizio di attività a carattere esclusivamente personale o domestico[209]. Con "rappresentante autorizzato" si intende una persona fisica o giuridica ubicata o stabilita nell'UE che ha ricevuto e accettato un mandato scritto da un fornitore di un sistema di IA o di un modello di IA per finalità generali al fine, rispettivamente, di adempiere ed eseguire per suo conto gli obblighi e le procedure stabiliti dal Regolamento. L'importatore è una persona

[209] *Regolamento UE 2016/67*, in Gazzetta Ufficiale dell'Unione Europea, art. 2 lett. c).

fisica o giuridica ubicata o stabilita nell'UE che immette sul mercato un sistema di IA recante il nome o il marchio di una persona fisica o giuridica stabilita in un Paese terzo. Il distributore è una persona fisica o giuridica che mette a disposizione un sistema di IA sul mercato dell'UE, diversa dal fornitore o dall'importatore.

L'art. 2 precisa, invece, l'ambito di applicazione territoriale del Regolamento, che come quello del GDPR, si rivolge agli operatori sia all'interno che all'esterno dell'UE, vincolando quindi anche gli operatori stabiliti in un paese terzo[210]. Ogni qualvolta un sistema di IA possa produrre degli effetti nel territorio dell'Unione, ad esso si applicano le disposizioni dell'AIA, indipendentemente dal fatto che gli sviluppatori, coloro che hanno immesso nel mercato il sistema o gli utilizzatori siano stabiliti in uno degli Stati membri: questo perché, come si è detto in precedenza, l'obiettivo del Regolamento è tutelare i diritti fondamentali delle persone presenti nel territorio dell'Unione, in virtù della prospettiva antropocentrica che ispira la normativa.

Oltre alla già menzionata *household exemption*, l'*AI Act* non si applica nei casi di sistemi di IA o modelli di IA sviluppati per il solo scopo di ricerca e sviluppo scientifici, in materia di politica sociale degli Stati membri, o ancora nel caso di sistemi con scopi militari, di difesa o di sicurezza nazionale.

Oltre alla posizione centrale che viene conferita all'essere umano, che è il soggetto destinatario della tutela, il secondo elemento strutturale del Regolamento, strettamente collegato alla tutela dei diritti fondamentali, è l'approccio c.d. "*risk based*". Il legislatore europeo volge la sua attenzione ai rischi potenziali dell'innovazione tecnologica per i cittadini dell'Unione, e solo attraverso questa lente di osservazione vengono disciplinati i sistemi di IA.

Prima della redazione della versione definitiva del Regolamento, tale approccio è stato criticato da chi riteneva che esso non avrebbe garantito un'adeguata salvaguardia dei diritti fondamentali. In particolare, nello *Statement*[211] sottoscritto da centoquattordici tra organizzazioni internazionali e associazioni

---

[210] Pappalardo, op. cit., pp. 34-35.

[211] Ci si riferisce qui a *An EU Artificial Intelligence Act for Fundamental Rights A Civil Society Statement*. Si tratta di una dichiarazione collettiva, pubblicata il 30 Novembre 2021.

non governative, i soggetti firmatari hanno rivolto il loro appello al Consiglio, al Parlamento e ai governi degli Stati membri, presentando nove obiettivi che avrebbero dovuto orientare la revisione della proposta di Regolamento.

Nel documento si è rilevato come la proposta, che operava unicamente una classificazione *ex ante* dei sistemi di IA in base al criterio del rischio, fosse "disfunzionale", in quanto "il livello di rischio dipende anche dal contesto in cui un sistema viene impiegato e non può essere completamente determinato in anticipo"[212]. Per garantire l'adeguatezza della risposta normativa dell'AIA, occorreva che la categorizzazione esistente fosse flessibile, e ciò poteva essere garantito solo dalla possibilità di integrare la classificazione sulla base delle ricadute, osservabili *ex post*, dell'impiego delle tecnologie IA sui diritti fondamentali[213].

La versione definitiva del testo, approvata dal Parlamento dell'Unione il 13 Giugno 2024 ed entrata ufficialmente in vigore a partire dal primo Agosto dello stesso anno, ha parzialmente recepito queste indicazioni.

L'articolo 27, - rubricato *Valutazione d'impatto sui diritti fondamentali per i sistemi di IA ad alto rischio*, - sancisce che, qualora si voglia utilizzare un'IA qualificabile come sistema "ad alto rischio", ai sensi dell'art. 6 e dell'*Annex III* del Regolamento, sarà necessario vagliare la sua potenzialità lesiva nei confronti di uno o più diritti fondamentali protetti dalla Carta dei diritti fondamentali dell'Unione Europea.

I rischi sono categorizzati in inaccettabile, - si parla, in questo senso di "pratiche vietate",- alto, limitato, minimo e nullo.

Il rischio inaccettabile è disciplinato all'art. 5 dell'*AI Act*. Dal 2 Febbraio 2025 è entrato in vigore il primo "scaglione applicativo" della normativa, che riguarda proprio le pratiche vietate: le pratiche, gli usi e gli impieghi della tecnologia di IA sono vietati ove si tratti di tecniche subliminali, manipolative o ingannevoli che abbiano un intento distorsivo del comportamento dell'individuo, e che ne influenzino la capacità decisionale, causando un danno significativo; nel caso in cui sfruttino la vulnerabilità di una persona fisica, dovuta all'età, alla disabilità

[212] Traduzione di Nardocci, op. cit., p. 270.
[213] Nardocci, op. cit., p. 271.

o ad una specifica situazione sociale o economica, per distorcerne il comportamento, causando, ancora, un danno significativo; nel caso in cui compiano attività di *social scoring*, ossia la valutazione o la classificazione di individui o gruppi in base al comportamento sociale o alle caratteristiche personali, causando un trattamento negativo o sfavorevole di tali persone; nel caso in cui compiano attività di valutazione del rischio che un individuo commetta reati solo sulla base del profilo o dei tratti e delle caratteristiche della personalità. Ancora, nel caso in cui i sistemi di IA effettuino la compilazione di database di riconoscimento facciale; nel caso in cui deducano le emozioni nei luoghi di lavoro o negli istituti scolastici, ad eccezione dei sistemi impiegati per motivi medici e di sicurezza; qualora si tratti di sistemi di categorizzazione biometrica che deducono caratteristiche "sensibili"; infine, nei casi di identificazione biometrica remota (RBI) "in tempo reale" in spazi accessibili al pubblico per le forze dell'ordine.

In sintesi, le otto pratiche vietate rispondono all'esigenza di tutela di quattro principi cardine: la libertà di scelta e l'autodeterminazione, il principio di non discriminazione, lo stato di diritto e la presunzione di innocenza, il rispetto della vita privata, e dunque la protezione dei dati personali[214].

In caso di violazione dell'art. 5, le sanzioni previste sono quelle ex art. 99: si tratta delle sanzioni pecuniarie del massimo ammontare, cioè 35 milioni di euro o 7% del fatturato totale annuo a livello mondiale dell'esercizio finanziario precedente.

I sistemi ad alto rischio sono disciplinati nel Capo III sezione I del Regolamento, agli artt. 6 e 7, nell'Allegato I e nell'Allegato III. Si tratta di una categoria di sistemi di IA molto ampia, la più rilevante, e con un forte impatto: i requisiti, gli adempimenti e i costi di *compliance* sono infatti particolarmente onerosi per le aziende i cui sistemi impiegati rientrino in tale gruppo[215]. All'interno dei sistemi di IA ad alto rischio sono previste due sottocategorie.

---

[214] G. Scafati, *Le pratiche vietate,* Navigare l'*European AI Act*, AIRIA Associazione per la Regolazione dell'Intelligenza Artificiale (a cura di), Wolters Kluwer, Milano 2024, p. 61.

[215] M. Colonna, *I sistemi ad alto rischio e i* General Purpose AI Models, Sezione I *I sistemi ad alto rischio,* Navigare l'*European AI Act*, AIRIA Associazione per la Regolazione dell'Intelligenza Artificiale (a cura di), Wolters Kluwer, Milano 2024, pp. 71-72.

Fanno parte del primo gruppo quei sistemi di IA ad alto rischio che soddisfano due condizioni cumulative. Innanzitutto, si deve trattare di sistemi che costituiscono componenti di sicurezza di un prodotto, o sistemi di IA che costituiscono essi stessi prodotti, disciplinati in una lista esaustiva contenuta nella normativa europea, in particolare nell'Allegato I[216]. Il Considerando 50 del Regolamento fa riferimento, tra gli altri, a prodotti quali macchine, giocattoli, ascensori, dispositivi medici. In secondo luogo, il prodotto o la componente di sicurezza del prodotto deve essere soggetto a una valutazione della conformità da parte di terzi ai fini dell'immissione sul mercato o della messa in servizio di tale prodotto[217].

Si tratta di settori che sono stati già ampiamente regolamentati dall'Unione, dunque le difficoltà maggiori risiedono nel processo di armonizzazione delle normative europee e di integrazione dei processi di compliance già presenti nelle aziende del settore; in questo senso il legislatore europeo ha attribuito un ruolo fondamentale alla Commissione, che dovrà garantire il raccordo tra i regolamenti unionali[218].

La seconda sottocategoria riguarda i sistemi di IA ad alto rischio definiti "indipendenti"[219]; si tratta di una categoria residuale, in cui rientrano tutti i sistemi di IA ad alto rischio non facenti parte della prima categoria. Essi sono indicati tassativamente nell'Allegato III; tuttavia è previsto un meccanismo di deroga, e l'attribuzione alla Commissione della facoltà di integrazione dei sistemi rientranti nella sottocategoria.

I settori interessati[220] sono la biometria, le infrastrutture critiche, l'istruzione e la formazione professionale, l'occupazione, la gestione dei lavoratori e l'accesso al lavoro autonomo, l'acceso a servizi privati essenziali e a prestazioni e servizi pubblici essenziali e fruizione degli stessi, l'attività di contrasto, la migrazione,

---

[216] *Regolamento (UE) 2024/1689*, in Gazzetta Ufficiale dell'Unione Europea, art. 6 par. 1 lett. a).
[217] *Regolamento (UE) 2024/1689*, in Gazzetta Ufficiale dell'Unione Europea, art. 6 par. 1 lett. b).
[218] Colonna, op. cit, pp. 72-73.
[219] *Regolamento (UE) 2024/1689*, in Gazzetta Ufficiale dell'Unione Europea, art. 6 par. 2.
[220] *Regolamento (UE) 2024/1689*, in Gazzetta Ufficiale dell'Unione Europea, Allegato III.

l'asilo e la gestione delle frontiere, l'amministrazione della giustizia e dei processi democratici.

Per quanto attiene alla deroga, il legislatore europeo ha previsto che un sistema con i requisiti di cui all'Allegato III non sia considerato ad alto rischio nel caso in cui "non presenta un rischio significativo di danno per la salute, la sicurezza o i diritti fondamentali delle persone fisiche, anche nel senso di non influenzare materialmente il risultato del processo decisionale"[221]. Il produttore deve quindi essere in grado di dimostrare che il sistema soddisfi almeno una delle quattro condizioni indicate[222]. La loro sussistenza permette che il sistema non venga classificato "ad alto rischio", pur rientrando nella casistica di cui all'Allegato III: esse attengono al carattere non decisivo o non completamente autonomo delle operazioni compiute dal sistema. Non sono ritenuti ad alto rischio, insomma, i sistemi di IA in cui è mantenuta una certa componente di partecipazione umana. Il terzo comma precisa poi che le attività di profilazione, invece, sono da considerarsi sempre "ad alto rischio"[223], secondo l'impostazione già adottata all'art. 22 del GDPR.

Una previsione particolarmente importante, allo stesso articolo 6[224], rimette all'utilizzatore finale la possibilità di utilizzare un sistema di IA ad alto rischio, qualora ritenga che non sia ad alto rischio: quando sia in grado, cioè, di provarne la "non lesività". Un'interpretazione unitaria dei paragrafi dell'art. 6, nella parte in cui ammettono deroghe alla qualificazione come "ad alto rischio", porta con sé delle perplessità sulla capacità di definizione di quali siano tali sistemi in modo tassativo ed univoco. Le valutazioni sul grado di lesività di ogni sistema di IA preso in considerazione rischiano dunque di variare notevolmente da caso a caso, con una conseguente frammentarietà sia nei rapporti tra gli Stati membri che all'interno dei singoli ordinamenti[225].

Sebbene solo il 10% dei sistemi di IA sia classificabile come ad alto rischio, una porzione rilevante delle fattispecie previste nell'Allegato III riguarda le

---

[221] *Regolamento (UE) 2024/1689*, in Gazzetta Ufficiale dell'Unione Europea, art. 6 par. 3 co. 1.
[222] *Regolamento (UE) 2024/1689*, in Gazzetta Ufficiale dell'Unione Europea, art. 6 par. 3 co. 2.
[223] *Regolamento (UE) 2024/1689*, in Gazzetta Ufficiale dell'Unione Europea, art. 6 par. 3 co. 3.
[224] *Regolamento (UE) 2024/1689*, in Gazzetta Ufficiale dell'Unione Europea, art. 6 par. 4.
[225] Nardocci, op. cit., pp. 281-282.

amministrazioni pubbliche. La ragione risiede nel fatto che l'impiego dell'IA nel settore pubblico può, più che in altri settori, riprodurre stereotipi e generare *bias*[226], come si è cercato di evidenziare nel corso di questa trattazione.

Se da una parte la scelta del legislatore di compiere un elenco tassativo è più garantista, poiché, trattandosi di una materia complessa, un'enumerazione che fosse solo esemplificativa avrebbe portato ad un eccessivo spazio di discrezionalità, dall'altro rischia di non essere in grado di stare al passo con il progresso tecnologico. Si è cercato di stemperare la rigidità dell'elencazione contenuta nell'Allegato III con l'attribuzione alla Commissione del ruolo di aggiornamento periodico del testo, - si tratta della facoltà di integrazione cui si è fatto cenno.

In un settore in continua evoluzione, come lo sviluppo di tecnologie di Intelligenza Artificiale, l'approccio antropocentrico adottato dall'AIA non può che comportare la necessità di meccanismi di riequilibro, e di cadenzata ricalibrazione della disciplina, affinché la tutela dei diritti fondamentali, - su cui si struttura l'intero Regolamento, - sia effettiva.

Nella stessa ottica di aggiornamento della disciplina, e specificatamente al fine di sostenere lo sviluppo di tecnologie conformi allo standard europeo, si prevede l'obbligo di sottoporre tutti i sistemi ad alto rischio ad una sorveglianza preventiva nell'ambito di spazi di sperimentazione normativa per l'IA[227]. Si tratta delle c.d. *regulatory sandbox,* strumenti sperimentali il cui utilizzo è garantito dalla previsione di clausole di sperimentazione che permettono alle autorità di applicare la legislazione con un certo grado di flessibilità, in modo da favorire l'innovazione e mitigare la rigidità del sistema. Questi spazi controllati permettono di addestrare sistemi di IA innovativi che siano conformi ai requisiti normativi, senza che gli ingenti investimenti vengano vanificati nell'elaborazione di prodotti e servizi che risultino poi non conformi alla disciplina unionale[228].

---

[226] V. D'Antino, *L'approccio basato sul rischio nell'*AI Act*: un nuovo paradigma di regolazione dell'intelligenza artificiale*, in federalismi.it, 2025, p. 21.
[227] *Regolamento (UE) 2024/1689*, in Gazzetta Ufficiale dell'Unione Europea, art. 57.
[228] D'Antino, op. cit., p. 25.

In ultima analisi, ciò che emerge dall'adozione del *risk-based approach,* è la ricerca di un punto di contemperamento tra il fenomeno economico dell'IA, - che deve essere incoraggiato e disciplinato, stabilendone le modalità di produzione, circolazione, utilizzo -, e i diritti della persona, la cui tutela deve essere garantita avverso i rischi che dall'impiego dell'IA possono derivare. Sulla base della valutazione circa il livello di gravità della minaccia per le persone fisiche, il Regolamento stabilisce le modalità di circolazione di quel prodotto, fino a prevedere, nei casi di rischio inaccettabile, delle pratiche vietate. Nell'impianto dell'*AI Act* l'illiceità colpisce a monte alcuni tipi di prodotti, impedendo che talune pratiche possano essere realizzate, nella prospettiva di un'affermazione assoluta dei diritti della persona. Pare coerente dunque la previsione per cui è precluso agli individui l'esercizio della facoltà di scelta circa l'utilizzazione di certi prodotti, attraverso una delineazione del perimetro mercato, con il risultato di circoscrivere l'attività di impresa[229].

L'*AI Act* è il primo testo normativo volto a regolare il fenomeno dell'Intelligenza Artificiale che sia stato compiutamente realizzato a livello globale: la scelta dell'Unione di servirsi dello strumento del regolamento per disciplinare la materia è espressiva della volontà di garantire un omogeneo assoggettamento di tutti gli Stati membri alle stesse disposizioni in materia, sebbene con le dovute precauzioni cui si è accennato sopra, legate alla difficoltà di ottenere interpretazioni uniformi nei diversi ordinamenti nazionali, per via dell'eccessiva indeterminatezza di alcune disposizioni. La vaghezza colpisce, - ed è uno degli aspetti più criticati del Regolamento, - la stessa nozione di Intelligenza Artificiale: pare problematica, infatti, la scelta del legislatore europeo di estenderne il significato fino a includervi semplici *software* o tecnologie di IA non *machine learning*, stabilendo anche per questi analoghi requisiti di *compliance*[230]. A ciò si aggiunge che, sebbene la decisione di disciplinare la

---

[229] V. Ricciuto, *La costruzione di un mercato antropocentrico dell'Intelligenza Artificiale tra diritto dell'Unione e diritto nazionale*, in A. C. Amato Mangiameli, P. Simone, C. Venturini (a cura di), Nuove tecnologie e diritto. Opportunità e rischi per la società del XXI secolo, Università degli Studi di Roma "Tor Vergata", Collana delle pubblicazioni del Dipartimento di Giurisprudenza. Quaderni del dottorato di ricerca in Diritto Pubblico, Wolters Kluwer, Milano 2025, pp. 87-88.

[230] Nardocci, op. cit., pp. 287-288.

materia con un atto regolamentare pare condivisibile se si guarda alla sua *ratio*, e cioè l'armonizzazione a livello unionale in una materia che è necessariamente transnazionale, pare è opinabile l'ampio spazio di manovra lasciato alla Commissione, che sembra non essere coerente con tale scelta di sistema.

Se, prediligendo il regolamento alla direttiva, si è voluta limitare la discrezionalità degli Stati membri, non appare giustificabile la scelta del legislatore europeo di demandare numerose decisioni ad un organo non rappresentativo come la Commissione, con ricadute sui principi dello stato di diritto, la legalità e la tutela dei diritti fondamentali[231]. È il caso, - ed è forse l'esempio più rilevante, - dell'art. 27, il quale riconosce alla Commissione un ampio margine di discrezionalità.

Interessante ai fini di questa trattazione è l'innovativa interpretazione dell'approccio basato sul rischio fornita dal Regolamento: si tratta della prospettiva *top down*, cioè "dall'alto". È il legislatore europeo, infatti, a fare le scelte più importanti in materia di IA: a definire i sistemi ad alto rischio, le pratiche vietate, le eccezioni al divieto di ricorso a specifici sistemi, attraverso norme di diritto positivo che, con classificazioni e definizioni astratte, disciplinano *ex ante* la materia[232].

Si tratta, dunque, di un'impostazione di sistema, che dovrà essere valutata sulla base delle scelte operative che ne conseguiranno e della capacità dell'atto di rispondere alle esigenze di relazione con gli ordinamenti extraeuropei, che si ispirano a principi diversi – come quello statunitense che, si vedrà, ha optato per una strategia antitetica, con l'adozione di un atto di *soft law* che adotta un approccio *market-based*.

4.2. L'*AI Action Plan* statunitense

L'*AI Action Plan* ha visto la luce nel corso della seconda Presidenza Trump, a Luglio del 2025. Esso è stato preceduto dall' *Executive Order 14179* ("*Removing Barriers to American Leadership in Artificial Intelligence*"), con cui non solo è

[231] Nardocci, op. cit., p. 290.
[232] Nardocci, op. cit., p. 286.

stato cancellato il precedente *Executive Order 14110* di Biden, ma si è anche imposto all'*Office of Science and Technology Policy* (OSTP) di consegnare entro centottanta giorni un piano operativo per l'IA. Il documento è stato firmato da Donald Trump nel Luglio del 25, e si inserisce in un quadro globale estremamente differenziato: interessante ai fini di questa trattazione è la diversità di approccio rispetto al corrispettivo dell'Unione Europea, basato sul rischio e sulla tutela dei diritti fondamentali.

L'AIAP è un documento di programmazione politica, che si pone come ambizioso obiettivo quello del dominio globale in materia di sviluppo e implementazione dell'Intelligenza Artificiale negli Stati Uniti. È chiaro dunque che l'impostazione nordamericana sia volta a favorire la competitività del Paese, specie rispetto alla Cina, individuata quale principale concorrente, ma anche con riferimento all'Unione Europea che, nella prospettiva statunitense, costituisce una fonte di ostacolo per via della intensa regolamentazione della materia[233]. L'AIAP individua, dunque, nella deregolamentazione un elemento cruciale per il raggiungimento della supremazia globale nel mercato dell'IA: la politica del *laissez faire* si contrappone con forza all'*AI Act*, un Regolamento che, si è visto, traccia il perimetro del mercato attraverso un'attenta classificazione dei sistemi di IA in base al rischio.

La diversità di approccio emerge sin dalla scelta dell'atto preposto a regolare la materia: nel caso dell'Unione Europea si tratta di un atto di *hard law*, mentre l'AIAP statunitense è un documento programmatico, che si fa portavoce della strategia dell'amministrazione Trump, con una forte connotazione politica e un taglio ingegneristico. Il piano infatti è stato redatto, tra gli altri, da personalità autorevoli della *Silicon Valley* e del Dipartimento di Difesa statunitense.

Il documento si articola in tre Pilastri, attorno ai quali sono elaborate più le policy federali, destinate alle diverse aree di intervento.

Il primo Pilastro, denominato "*Accelerate AI Innovation*", si realizza innanzitutto con la rimozione degli ostacoli di ordine burocratico che possono

---

[233] E. Falletti, *La nuova "gold rush" dell'intelligenza artificiale: l'AI Action Plan negli Stati Uniti della seconda presidenza Trump*, Rivista Interdisciplinare sul Diritto delle Amministrazioni Pubbliche, CERIDAP, Fascicolo 1/2026, Gennaio – Marzo 2026, p. 144.

rallentare la crescita nel settore. In particolare, la scelta operata dall'AIAP va nella direzione di aumentare la sorveglianza federale, per evitare che i fondi federali vengano impiegati in Stati dove l'eccessiva regolamentazione in materia di IA porterebbe ad una loro dispersione. Non si tratta, dunque, di un tentativo di centralizzazione della legislazione di settore, ma neanche di un completo rinvio alla regolamentazione statale, con una riduzione della sorveglianza federale ai minimi livelli. L'opzione che l'amministrazione Trump pare accogliere è di compromesso, principalmente volta ad evitare che l'attività normativa statale contraddica l'impostazione del governo: l'erogazione di fondi federali viene dunque subordinata all'effettiva *deregulation* posta in essere nei singoli Stati, attraverso un'indagine sommaria dei provvedimenti statali più significativi in ambito IA[234].

Si tratta, a ben vedere, di una deregolamentazione *ex post,* che ha le sue radici nella cultura statunitense del *try first,* e che risulta coerente con la matrice giuridica di *common law*: la valutazione giudiziaria non può che operare solo dopo l'introduzione di nuove tecnologie, poiché si fonda sul precedente vincolante[235]. È chiaro il tentativo di deresponsabilizzare le *Big Tech,* soprattutto in riferimento alla produzione di beni destinati ai consumatori: in questo senso, la distanza rispetto all'AIA europeo è enorme, in quanto quest'ultimo si fonda sull'opposto principio di garantire la massima tutela ai consumatori, attraverso quella che, agli occhi dell'amministrazione Trump, è certamente una "*over regulation*".

Numerose sono le policy indirizzate alla vigilanza sulla deregolamentazione negli Stati federati: tra queste, l'indagine, - denominata "*Request for Information*", rivolta alle imprese e al pubblico in genere -, sulla normativa federale che ostacoli l'innovazione, e di conseguenza adotti le azioni necessarie a fini deregolativi, condotta dall'*Office of Science and Technology Policy* (OSTP). Ancora, l'*Office of Management and Budget* (OMB), insieme alle

[234] Falletti, op. cit., p. 148.
[235] Falletti, op. cit., p. 149.

agenzie federali, deve compiere un'operazione di revisione o di cancellazione di normative che ostacolino inutilmente lo sviluppo dell'IA o il suo impiego[236].
Rilevante ai fini di questa trattazione è la revisione, disposta dal documento, dell'*AI Risk Management Framework* del *National Institute of Standards and Technology* (NIST). Il NIST ha un ruolo fondamentale nella governance IA statunitense: si tratta di un'agenzia del Dipartimento del Commercio degli Stati Uniti che dagli inizi del Novecento si occupa di innovazione al fine di aumentare la competitività industriale. Il Quadro di Gestione del Rischio dell'IA (AI RMF) mira ad essere una risorsa nella gestione dei rischi legati all'IA[237], e rappresenta, - pur mantenendo la stessa prospettiva di incoraggiamento all'autoregolamentazione delle aziende, - lo standard ufficiale in materia di rischio in ambito IA. Non è privo di rilevanza, dunque, che l'AIAP disponga che vengano eliminati dal documento i riferimenti alla "*misinformation, Diversity, Equity, and Inclusion, and climate change*"[238]. Con questa disposizione il governo esprime la necessità che la disinformazione, i criteri DEI (Diversità, Eguaglianza, Inclusione) e il cambiamento climatico non vengano inclusi tra i parametri delle valutazioni sulle aree di rischio dei sistemi di IA.
Si ravvisa un'ulteriore differenza significativa con il Regolamento europeo: nel caso dell'*AI Act,* il pilastro su cui poggia l'intero atto è proprio la garanzia che i sistemi di IA siano conformi ai valori unionali, in una prospettiva di protezione della salute, della sicurezza e dei diritti fondamentali, - indipendentemente da dove la tecnologia sia sviluppata[239], - mentre nel corrispettivo statunitense ciò che deve essere salvaguardato nella costruzione e nello sviluppo dei sistemi di IA è il *freedom of speech*. Esso è garantito nel Primo Emendamento della Costituzione statunitense, ed è da intendersi, nell'interpretazione originalista adottata nell'AIAP, come garanzia di libertà di espressione del pensiero.

---

[236] *Winning the Race: America's AI Action Plan*, White House, Luglio 2025, p. 3.
[237] M. Colombo, *La regolazione dell'IA: scenari internazionali*, *La normativa negli Stati Uniti D'America,* Sezione II, Navigare l'*European AI Act*, AIRIA Associazione per la Regolazione dell'Intelligenza Artificiale (a cura di), Wolters Kluwer, Milano 2024, p. 19.
[238] *AIAP*, cit., p. 4.
[239] E. Raffiotta, *I profili definitori ed i principi dell'*AI Act, Navigare l'*European AI Act*, AIRIA Associazione per la Regolazione dell'Intelligenza Artificiale (a cura di), Wolters Kluwer, Milano 2024, p. 47.

Tuttavia, pur ponendosi formalmente in una posizione di "apertura dello spazio informativo", propugnando il diritto all'autodeterminazione delle modalità di manifestazione del pensiero, di fatto il documento compie una "chiusura selettiva verso ciò che non è completamente aderente alla dottrina MAGA"[240]. Emerge chiaramente il forte carattere politico del documento, che pare attuare una censura delle rivendicazioni egalitarie, di accettazione della diversità, dell'uguaglianza e dell'inclusione, ritenute espressive dell'ideologia *woke*[241].
Modificando lo standard tecnico federale rappresentato dall'AI RMF, la Casa Bianca compie una regolamentazione innanzitutto politica, prima ancora che giuridica: secondo tale impostazione sono da scongiurare, innanzitutto, i *bias* ideologici. Si muove in questa stessa direzione la seconda raccomandazione di policy volta all'aggiornamento delle linee guida sugli appalti federali, affinché il governo stipuli contratti unicamente con gli sviluppatori LLM "di frontiera" in grado di garantire sistemi dichiaratamente oggettivi e privi di pregiudizi ideologici di matrice politica o culturale[242]. I modelli LLM fanno parte dell'IA generativa: si basano su un processo di apprendimento automatico e sono sistemi linguistici di grandi dimensioni[243]. Pare discutibile, dunque, la scelta di demandare a tali sistemi la tutela della libertà di espressione e dei diritti fondamentali, considerati i rischi che siffatti sistemi portano con sé, primo fra tutti il fenomeno delle *black box.* Anche con riferimento a quest'ultima raccomandazione, evidentemente, l'*AI Action Plan* pare andare in direzione opposta rispetto alle intenzioni espresse nell'*AI Act*: l'affidamento ai modelli di LLM che siano scevri da *bias* ideologici, - così come definiti dal documento, - della tutela del *freedom of speech,* è in netta collisione con l'approccio dell'Unione Europea che guarda ai sistemi di *machine learning* con preoccupazione, per via della opacità dei processi, e che per questo motivo ha sancito numerosi obblighi di trasparenza sulle modalità di funzionamento di detti sistemi.

---

[240] Falletti, op. cit., p. 153.
[241] Ibid.
[242] *AIAP,* cit., p. 4.
[243] Falletti, op. cit., p. 154.

Il documento prosegue indicando altre aree di intervento relative al Primo Pilastro: la promozione dei modelli di *Open-Source* per l'adozione dell'IA, il potenziamento delle competenze IA dei lavoratori americani, il rinnovamento della produzione e lo sviluppo scientifico in materia di IA e il contrasto dei "*synthetic media*" nell'ordinamento giuridico statunitense.

Si ravvisa, con riferimento a quest'ultimo ambito, un allontanamento momentaneo alla *deregulation* di cui è permeato l'intero atto, e un avvicinamento, seppur minimo, al Regolamento unionale. Con l'espressione "*synthetic media*" si intendono contenuti digitali come immagini, video o registrazioni audio prodotti con IA, così realistici da sembrare reali. In particolare, richiamando la legge federale "*TAKE IT DOWN Act*"[244], l'AIAP intende contrastare l'uso e la diffusione dei *deepfake*, e nello specifico proteggere le parti vulnerabili dalla produzione artificiale o dalla manipolazione di immagini, specialmente nel caso di *deepfake* sessualmente espliciti e non consensuali[245].

Il Secondo Pilastro è denominato "*Build American AI Infrastructure*". La supremazia in ambito IA non può essere ottenuta senza ingenti investimenti in settori quali l'energia, i semiconduttori e le infrastrutture dei data center. Anche qui, si evidenzia la necessità di una deregolamentazione, che deve avvenire, ad esempio, con una semplificazione delle procedure per data center e infrastrutture energetiche. Al fine di migliorare il rapporto tra fabbisogno energetico e sviluppo di IA, l'AIAP suggerisce di delegare le valutazioni ambientali a sistemi di IA, come ad esempio PermitAI, per migliorare la qualità delle analisi[246].

Si tratta di una serie di raccomandazioni volte all'implemento delle infrastrutture in ambito IA: lo sviluppo di una rete elettrica adeguata alla domanda, la formazione della forza lavoro statunitense, il miglioramento della capacità federale di risposta agli incidenti IA, il rafforzamento della *cybersecurity* attraverso la costruzione di data center ad alta sicurezza per uso militare e

---

[244] Si tratta del "*Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks Act*", una legge federale approvaya dal 119° Congresso [119th Congress Public Law 12] e firmata dal Presidente Trump il 19 Maggio 2025.
[245] Falletti, op. cit., p. 165.
[246] Falletti, op. cit., p. 168.

Intelligence[247]. Con riferimento al rapporto tra Intelligenza Artificiale e forza lavoro, viene sottolineata la necessità di colmare la mancanza di figure professionali altamente qualificate attraverso l'identificazione delle occupazioni essenziali e la definizione di quadri di competenze, utili per la progettazione di curricula e certificazioni[248]. Emerge anche qui la netta predominanza dell'approccio competitivo e industriale, quale *fil rouge* del documento.

Il Terzo Pilastro si riferisce al "*Lead in International AI Diplomacy and Security*". Per primeggiare a livello globale, "*America must do more than promote AI within its own borders*"[249]. In particolare, le raccomandazioni di policy per la realizzazione del *Third Pillar* sono: l'esportazione del *tech stack*[250] statunitense agli Alleati, l'arginamento dell'influenza cinese negli organismi internazionali, il rafforzamento dei controlli sull'esportazione di *compute* avanzato[251], l'eliminazione delle criticità nei controlli sull'esportazione di tecnologie per la produzione di chip, l'allineamento globale delle misure di protezione, la valutazione dei rischi di sicurezza nazionale nei *frontier models* e gli investimenti in *biosecurity*.

L'AIAP, dunque, traccia una *roadmap* che dovrà guidare gli Stati Uniti nell'evoluzione tecnologica legata all'IA; i tre Pilastri del documento individuano nella *deregulation*, nell'incremento delle infrastrutture IA e nella diplomazia, le priorità e il posizionamento strategico della Presidenza Trump[252]. Sono molteplici le perplessità legate alle scelte di policy nelle aree di intervento individuate dall'atto.

Nel primo Pilastro, emerge l'impostazione di sistema del modello statunitense, che si basa sulla deregolamentazione. Esso mira a creare una zona franca che possa assicurare l'egemonia globale e, in particolare, l'affermazione sulla potenza cinese – che, al contrario, ha adottato un modello di regolamentazione ancor più ferreo di quello dell'Unione Europea. Per perseguire questo scopo,

---

[247] *AIAP,* cit., pp. 14-19.
[248] Falletti, op. cit., p. 169.
[249] *AIAP*, cit., p. 20.
[250] È l'insieme di linguaggi, database, framework e server usati per creare un sito o un'app.
[251] *AIAP,* cit., p. 21.
[252] D. de Borba Santos Jùnior, R. A. Brinkhues, *America's AI action plan and global governance: convergences, divergences, and theoretical implications*, Brazilian Journal of Public Administraion, Rio de Janeiro 2025, p. 13.

però, l'amministrazione Trump sembra sottovalutare la necessità di una regolamentazione basata sul rischio, proporzionata, che dia rilevanza all'*accountability*, alla trasparenza e ai diritti fondamentali, quali condizioni indispensabili per un'innovazione che sia socialmente legittima[253].

La tutela degli individui, intesa come protezione delle categorie deboli da *bias* discriminatori perpetrati da sistemi di *machine learning* le cui modalità di funzionamento sono sconosciute, non pare destare negli Stati Uniti la medesima preoccupazione suscitata in ambito unionale. L'*AI Act,* si è visto, si muove in una prospettiva di estrema sensibilità rispetto alle nuove minacce che l'epoca digitale rivolge ai diritti fondamentali, primi fra tutti la dignità umana e l'uguaglianza: l'intera struttura del Regolamento si fonda sulla valutazione dell'impatto che i sistemi di IA hanno sugli individui. L'*AI Action Plan,* al contrario, eleva il *freedom of speech* quale principio guida[254], in nome del quale altri principi possono essere sacrificati – si pensi alla disposizione che prevede l'eliminazione dei criteri DEI dall'AI RMF. La deresponsabilizzazione per i colossi globali dell'IA che ne consegue non può che apparire problematica, specialmente se si considera che gli algoritmi vengono impiegati in ambiti particolarmente sensibili quali la giustizia, la sicurezza, e la sanità, e che quindi una vigilanza sugli stessi è resa necessaria dalle conseguenze che un loro impiego non regolamentato avrebbe sull'individuo e sulla collettività.

L'importanza che viene data all'implemento delle infrastrutture in ambito IA – il secondo Pilastro del documento, - è, anch'essa, una misura innanzitutto politica, convincente e allineata all'AIA europeo nella parte in cui adotta un approccio *secure-by-design*, previsto anche nell'Unione per i sistemi ad alto rischio. Tuttavia, perplessità sorgono se si considera che non viene fatta menzione dell'impatto che le grandi infrastrutture IA hanno sull'ambiente e sulla sostenibilità. Ciò non sorprende, se si considera che tra i parametri di valutazione del rischio di cui si è chiesta l'eliminazione è incluso il riferimento al "*climate change*": l'esclusione dell'agenda climatica dal documento è senza dubbio una scelta di posizionamento politico, che sebbene sia finalizzata ad aumentare la

[253] Ibid.
[254] Ibid.

competitività a livello globale, di fatto rischia di minare il riconoscimento internazionale del modello americano[255].

Per quanto riguarda il Terzo Pilastro, è qui che emergono con più intensità le policy geopolitiche e le preoccupazioni legate alla competitività della Cina. Viene sancito un principio di sicurezza tecnologica, per cui la diffusione di tecnologie di origine nazionale deve avvenire esclusivamente verso Paesi alleati e Partner affidabili. È la parte del documento più strettamente attinente alla diplomazia: è chiara l'esigenza di un quadro regolatorio condiviso, e in questo senso gli Stati Uniti intendono promuovere un'armonizzazione progressiva dei regimi di controllo tecnologico, al fine di "evitare fenomeni di *backfill* tecnologico, ossia la fornitura indiretta di tecnologie sensibili da parte di partner non vincolati da restrizioni equivalenti"[256]. L'approccio sovranista, che emerge con forza con il terzo Pilastro, ha nell'esportazione controllata dell'innovazione, nell'allineamento regolatorio e nella diplomazia della sicurezza i suoi punti cardine.

Sono rari, dunque, i punti di contatto con l'Unione Europea, che con l'*AI Act* ha fatto una scelta di regolamentazione *ex ante,* di normativizzazione del fenomeno e di classificazione preventiva del rischio, incardinando l'intero Regolamento sulla tutela dei diritti fondamentali. Se nel caso dell'AIA l'intento di uniformare a rigidi standard la normativa degli Stati membri è ravvisabile sin dallo strumento normativo utilizzato, cioè l'atto regolamentare, allo stesso modo la tendenza statunitense ad evitare regolamentazioni che possano ostacolare la crescita economica e lo sviluppo tecnologico sono alla base dell'emanazione dell'*AI Action Plan*, un documento di policy che mira ad orientare le legislazioni degli Stati federati e i finanziamenti, ma che non è immediatamente applicabile. Tale diversità di intenti si ravvisa anche nell'ambito di applicazione: l'*AI Act* si applica a tutti i sistemi di IA, che vengono classificati non in base al settore di appartenenza ma in base al livello di rischio, ivi compresi i sistemi immessi nel mercato europeo da soggetti terzi, mentre il corrispondente atto americano compie una distinzione su base settoriale, facendo riferimento a *framework*

[255] de Borba Santos Jùnior, Brinkhues, op. cit., p. 14.
[256] Falletti, op. cit., p. 171.

tecnici e a regolazioni di settore. L'AIAP non dispone veri e propri divieti, piuttosto mira a disincentivare la regolazione in ambito IA da parte degli Stati federati, mentre l'AIA sancisce le pratiche vietate e rigidi requisiti per i sistemi ad alto rischio, prevedendo ingenti sanzioni pecuniarie.

Si ravvisa, in ogni caso, la consapevolezza in entrambi i documenti della necessità di standard uniformi a livello internazionale, sebbene ad oggi paia difficile ipotizzare una convergenza tra approcci così differenti, - e in alcuni casi antitetici, - a livello globale.

Bibliografia

I. M. Alagna, *Diritto all'oblio e mancato oscuramento dei dati sensibili: responsabilità, profili sanzionatori e rimedi pratici applicabili*, Giuffrè, Milano 2023.

S. Amato, *Inquietudini digitali. Il "black box effect"*, "Rivista di filosofia del diritto", Fascicolo 1, giugno 2025, in "Il Mulino – Rivisteweb", Bologna 2025.

G. Avanzini, *Intelligenza artificiale,* machine learning *e istruttoria procedimentale: vantaggi, limiti ed esigenze di una specifica* data governance, in A. Pajno, F. Donati, A. Perrucci (a cura di), Intelligenza artificiale e diritto: una rivoluzione?, Volume 2, Amministrazione, responsabilità, giurisdizione, Asrtrid, Il Mulino, Bologna 2022.

C. Barbaro, *Uso dell'intelligenza artificiale nei sistemi giudiziari: verso la definizione di principi etici condivisi a livello europeo? I lavori in corso alla Commissione europea per l'efficacia della giustizia (Cepej) del Consiglio d'Europa*, Obiettivo 1. Una giustizia (im)prevedibile, in "Questione Giustizia", n. 4/2018, Magistratura Democratica 2018.

P. Barile, *La Costituzione come norma giuridica: profilo sistematico*, Barbera Editore, Firenze 1951.

F. Basile, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, "Diritto penale e uomo – Criminal Law and Human Condition", fascicolo n.10/2019, Diritto penale contemporaneo 2019.

F. Basile, *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, in F. Basile, M. Caterini, S. Romano (a cura di), Il sistema penale ai confini delle *hard sciences*: Percorsi epistemologici tra neuroscienze e intelligenza artificiale, Pacini Giuridica 2021.

C. Benetazzo, *IA, giustizia e P.A.: la sfida etica ed antropologica*, in federalismi.it, 12 Marzo 2025.

C. Bernardo, *Il principio di uguaglianza*, in L. Delli Priscoli (a cura di), La Costituzione vivente, Giuffrè, Milano 2023.

R. Bin, G. Pitruzzella, *Diritto Costituzionale*, G. Giappichelli Editore, Torino 2020.

A. Buratti, *Western Costitutionalism. History, Institutions, Comparative Law,* Springer e G. Giappichelli Editore 2023.

A. Carboni, *Giustizia algoritmica e applicazione dei precedenti parlamentari*, in D. De Lungo, G. Rizzoni (a cura di), Le assemblee rappresentative nell'era dell'intelligenza artificiale: profili costituzionali, G. Giappichelli Editore, Torino 2025.

R. Cavallo Perin, *Ragionando come se la digitalizzazione fosse data*, in Diritto amministrativo, Rivista trimestrale, 2/2020.

M. Colombo, *La regolazione dell'IA: scenari internazionali*, *La normativa negli Stati Uniti D'America,* Sezione II, Navigare l'*European AI Act*, AIRIA Associazione per la Regolazione dell'Intelligenza Artificiale (a cura di), Wolters Kluwer, Milano 2024.

M. Colonna, *I sistemi ad alto rischio e i* General Purpose AI Models, Sezione I *I sistemi ad alto rischio,* Navigare l'*European AI Act*, AIRIA Associazione per la Regolazione dell'Intelligenza Artificiale (a cura di), Wolters Kluwer, Milano 2024.

G. D'Acquisto, *Decisioni algoritmiche: equità, causalità, trasparenza*, G. Giappichelli Editore, Torino 2022.

F. Dal Monte, *The philosophy of the GDPR*, in E. Longo, A. Pin, F. Viglione (a cura di), Data protection in context: between privacy and AI, Giuffrè, Milano 2025.

V. D'Antino, *L'approccio basato sul rischio nell'*AI Act*: un nuovo paradigma di regolazione dell'intelligenza artificiale*, in federalismi.it, 2025.

D. de Borba Santos Jùnior, R. A. Brinkhues, *America's AI action plan and global governance: convergences, divergences, and theoretical implications*, Brazilian Journal of Public Administraion, Rio de Janeiro 2025.

G. C. di San Luca, M. Paladino, *Decisione amministrativa e intelligenza artificiale,* in federalismi.it, 29 gennaio 2025.

P. Falletta, *Il* freedom of information act *italiano e i rischi della trasparenza digitale*, in federalismi.it, n. 23/2016.

E. Falletti, *Discriminazione algoritmica: una prospettiva comparata*, G. Giappichelli Editore, Torino 2022.

E. Falletti, *La nuova "gold rush" dell'intelligenza artificiale: l'AI Action Plan negli Stati Uniti della seconda presidenza Trump*, Rivista Interdisciplinare sul Diritto delle Amministrazioni Pubbliche, CERIDAP, Fascicolo 1/2026, Gennaio – Marzo 2026.

M. Fasan, *Intelligenza artificiale e costituzionalismo contemporaneo: principi, diritti e modelli in prospettiva comparata*, Collana della Facoltà di Giurisprudenza dell'Università di Trento, 49, Università di Trento 2024.

M. Fasan, *L'intelligenza artificiale alla prova dell'eguaglianza. L'evoluzione del diritto costituzionale tra scenari attuali e futuri*, in M. Leone (a cura di), Annali di studi religiosi, "Aracne", Prima edizione, 26/2025.

M. Fasan, *I principi costituzionali nella disciplina dell'Intelligenza Artificiale. Nuove prospettive interpretative*, in C. Casonato, M. Fasan, S. Penasa (a cura di), Diritto e Intelligenza Artificiale, "Sezione monografica – DPCE ONLINE", Fascicolo n. 1/2022.

M. Fioravanti, *Il principio di eguaglianza nella storia del costituzionalismo moderno*, in "Il Mulino", a. II (1999).

L. Floridi, *Etica dell'intelligenza artificiale: sviluppi, opportunità, sfide*, Raffaello Cortina Editore, Milano 2022.

M. Gialuz, *Quando la giustizia penale incontra l'Intelligenza Artificiale: luci e ombre dei* risk assessment tools *tra Stati Uniti ed Europa*, "Diritto Penale Contemporaneo".

P. Gori, *I diritti fondamentali*, in L. Delli Priscoli (a cura di), La Costituzione vivente, Giuffrè, Milano 2023.

E. Hine, C. Novelli, M. Taddeo, L. Floridi, *Supporting Trustworthy AI Through Machine Unlearning*, Science and Engineering Ethics, National Library of Medicine, Springer 2024.

E. Longo, *From data protection to AI regulation*, in E. Longo, A. Pin, F. Viglione (a cura di), Data protection in context: between privacy and AI, Giuffrè, Milano 2025.

P. G. Lucifredi, *Almagesto costituzionale*, Giuffrè Editore, Milano 1998.

M. Manetti, *Pluralismo dell'informazione e libertà di scelta*, in "AIC – Associazione Italiana dei Costituzionalisti", 1 (2012).

B. Marchetti, *La garanzia dello* human in the loop *alla prova della decisione amministrativa algoritmica*, in BioLaw Journal – Rivista di Biodiritto, 2/2021.

D. Messina, *La tutela della dignità nell'era digitale: prospettive e insidie tra protezione dei dati, diritto all'oblio e Intelligenza Artificiale*, Editoriale Scientifica, Napoli 2023.

E. Mollick, *Co-intelligence*, Londra, WH Allen, 2024.

C. Nardocci, *Algoritmi, eguaglianza, discriminazione: le sfide dell'intelligenza artificiale*, G. Giappichelli Editore, Torino 2025.

C. Nardocci, *Dalla "self-regulation" alla frammentata regolamentazione dei sistemi di intelligenza artificiale: uno sguardo alla (diversa) prospettiva statunitense*, "Diritto pubblico comparato ed europeo", Fascicolo 4, ottobre-dicembre 2024, in "Il Mulino – Rivisteweb", Bologna 2024.

A. Papa, *La problematica tutela del diritto all'autodeterminazione informativa nella* big data society, in *Liber amicorum* per Pasquale Costanzo – Diritto costituzionale in trasformazione, vol. I – Costituzionalismo, Reti e Intelligenza Artificiale, Collana di studi di *Consulta OnLine*, 3, Genova 2020.

M. Pappalardo, *Navigare l'European AI Act*, AIRIA Associazione per la Regolazione dell'Intelligenza Artificiale (a cura di), Wolters Kluwer, Milano 2024.

C. Parodi, V. Sellaroli, *Sistema penale e intelligenza artificiale: molte speranze e qualche equivoco*, in "Diritto Penale Contemporaneo", fascicolo n. 6/2019.

F. Patroni Griffi, *La trasparenza della pubblica amministrazione tra accessibilità totale e riservatezza*, in federalismi.it n. 8/2013, Roma 2013.

R. Perona, *Giustizia artificiale intelligente? Spunti per il costituzionalista europeo a partire dall'esperienza colombiana*, in federalismi.it, n. 26/2025.

F. Polacchini, *Il principio di eguaglianza*, in L. Mezzetti (a cura di), Principî costituzionali, G. Giappichelli Editore, Torino 2011.

B. Ponti, *Il codice della trasparenza amministrativa: non solo riordino, ma ridefinizione complessiva del regime della trasparenza amministrativa online,* in www.neldiritto.it, 2013.

A.E.R. Prince, D. Schwarcz, *Proxy discrimination in the age of artificial intelligence and big data*, in *Iowa Law Review,* 2020.

E. Raffiotta, *I profili definitori ed i principi dell'*AI Act, Navigare l'*European AI Act*, AIRIA Associazione per la Regolazione dell'Intelligenza Artificiale (a cura di), Wolters Kluwer, Milano 2024.

V. Ricciuto, *La costruzione di un mercato antropocentrico dell'Intelligenza Artificiale tra diritto dell'Unione e diritto nazionale*, in A. C. Amato Mangiameli, P. Simone, C. Venturini (a cura di), Nuove tecnologie e diritto. Opportunità e rischi per la società del XXI secolo, Università degli Studi di Roma "Tor Vergata", Collana delle pubblicazioni del Dipartimento di Giurisprudenza. Quaderni del dottorato di ricerca in Diritto Pubblico, Wolters Kluwer, Milano 2025.

G. Scafati, *Le pratiche vietate,* Navigare l'*European AI Act*, AIRIA Associazione per la Regolazione dell'Intelligenza Artificiale (a cura di), Wolters Kluwer, Milano 2024.

G. Sgueo, *La funzione pubblica sintetica. Tre domande su intelligenza artificiale generativa e pubblica amministrazione*, in Rivista Trimestrale di Diritto Pubblico, num. 4/2024.

A. Simoncini, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in "BioLaw Journal - Rivista di BioDiritto", 1/19, Trento, Febbraio 2019.

A. Simoncini, *Amministrazione digitale algoritmica. Il quadro costituzionale*, in R. Cavallo Perin, D. U. Galetta (a cura di), Il diritto dell'amministrazione pubblica digitale, Giappichelli Editore, Torino 2020.

C. Tripodina, *L'argomento originalista nella giurisprudenza costituzionale in materia di diritti fondamentali*, in F. Giuffrè e I. Nicotra (a cura di), Lavori preparatori ed *original intent* nella giurisprudenza della Corte costituzionale, G. Giappichelli Editore, Torino 2008.

S. Talamo, *La trasparenza totale. Partenza lenta ma grandi orizzonti*, Smart24 PA+, Gruppo24Ore, Milano 2019.

A. Vedaschi, *Il principio personalista*, in L. Mezzetti (a cura di), Principî costituzionali, G. Giappichelli Editore, Torino 2011.

L. Viola, *Interpretazione della legge con modelli matematici. Processo, a.d.r., giustizia predittiva*, Volume I, Seconda edizione, Centro Studi Diritto Avanzato Edizioni, Milano 2018.

G. Zagrebelsky, *Trent'anni dopo. Sulle riletture del* Diritto mite, in "il Mulino", Bologna, a. XLV, 2025.

D. Zingales, Risk assessment*: una nuova sfida per la giustizia penale?*, "Diritto penale e uomo – Criminal Law and Human Condition", fascicolo n. 12/2021.

S. Zuboff, *Il capitalismo della sorveglianza: il futuro dell'umanità nell'era dei nuovi poteri*, Luiss University Press, Roma 2019.

Siti web

www.garanteprivacy.it

www.neldiritto.it

www.semanticscholar.org