Alessandra Langella(*)

ASPETTI ETICI DELL'INTELLIGENZA ARTIFICIALE

Abstract: The essay considers the ethical aspects related to the use of artificial intelligence. European sources that invoke the ethics of AI are analyzed, as well as the legal Documents that seek to define an appropriate ethical and legal framework, based on the values of the Union, to calibrate the effects that the digital revolution deploys in the various and numerous contexts in which AI is used to ensure an anthropocentric approach to the technology that does not harm the fundamental interests protected by the legal system.

1. — *Etica e IA.*

L'uso dell'intelligenza artificiale (IA)(1) è sempre più diffuso in moltissimi settori della vita quotidiana (attività di lavoro, pratiche mediche, istruzione, giustizia, ricerca, politiche economiche e sociali).

L'IA costituisce uno strumento di grande opportunità ma il suo sviluppo ed utilizzo se non correttamente gestito può rappresentare anche numerose minacce per le libertà degli individui.

In questo senso, il funzionamento della nuova tecnologia solleva numerose questioni di natura etica e giuridica.

(*) Università degli Studi di Perugia.

(1) La Proposta di Regolamento del Parlamento europeo e del consiglio che stabilisce regole armonizzate sull'intelligenza artificiale e modifica alcuni atti legislativi dell'Unione definisce (art. 3, (1)) quale sistema di intelligenza artificiale «un software (sviluppato con una o più delle tecniche e degli approcci elencati nell'allegato I) che può, per una determinata serie di obiettivi definiti dall'uomo, generare output quali contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono».

Per assicurare che i sistemi di intelligenza artificiale risultino coerenti con i diritti e gli interessi fondamentali dei singoli, oltre che capaci di portare benefici validi per l'intera collettività è fondamentale sviluppare un quadro di regole specifiche.

La legislazione, sia a livello nazionale che a livello europeo, tenta di affrontare le questioni derivanti dall'uso dell'intelligenza artificiale attraverso la previsione di un quadro sorretto dai valori etici generali per un approccio antropocentrico e filantropico.

Del resto, il rispetto di requisiti etici è un presupposto indefettibile per lo sviluppo di un'IA sicura, giusta ed affidabile sia per chi la sviluppa e distribuisce, sia per chi ne fa uso.

In questo senso, un'IA etica sarebbe una «proposta vantaggiosa per tutti[2]».

Il richiamo all'etica dell'IA è compiuto in molti documenti europei relativi alla materia.

La Comunicazione della Commissione del 25 aprile 2018 su "L'intelligenza artificiale per l'Europa" (COM (2018) 237 final) indica come necessario in tema di IA «assicurare un quadro etico e giuridico adeguato basato sui valori dell'Unione e coerente con la Carta dei diritti fondamentali»; e nel Libro bianco sull'intelligenza artificiale del 19 febbraio 2020 (COM(2020) 65) la Commissione afferma nelle conclusioni che l'IA è una «tecnologia strategica che può offrire molti benefici ai cittadini, alle imprese e alla società nel suo insieme, a condizione che segua un approccio antropocentrico, etico, sostenibile e rispettoso dei valori e dei diritti fondamentali».

Allo stesso modo, la Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione concernenti il quadro relativo agli aspetti etici dell'intelligenza artificiale della robotica e delle tecnologie correlate (2020/2012(INL)) auspica (punto 9) l'integrazione di un solido approccio etico attraverso una Proposta di Regolamento sui principi etici per lo sviluppo, la diffusione e l'utilizzo dell'intelligenza artificiale, della robotica e delle tecnologie correlate; ed è quanto avviene

---

[2] Comunicazione della Commissione *Creare fiducia nell'intelligenza artificiale incentrata sull'uomo* (COM (2019) 168 final).

nella Proposta di Regolamento del 2021 che stabilisce regole armonizzate sull'IA ((2021) 206 final).

2. — *Le componenti dell'etica applicate all'IA.*

Il fenomeno dell'applicazione dell'IA deve rispettare tutte le leggi e i regolamenti applicabili (componente della legalità) i quali, come accennato, prevedono la conformità dei sistemi di IA anche ai principi etici (componente dell'eticità) [3].

Il concetto di etico è "orientato" nello spazio e nel tempo. Esso può assumere significati molto diversi e persino contrastanti, e a fronte di questo relativismo, si deve garantire che i principi dell'etica aderiscano e restino stabilmente ancorati ai diritti fondamentali.

Come indicato, ad esempio, nella Risoluzione del Parlamento Europeo del 2020 sopra citata, l'IA può dirsi etica quando rispetta la dignità umana, l'autonomia e l'autodeterminazione dell'individuo; se garantisce la parità di trattamento e l'assenza di discriminazioni per tutti, comprese le minoranze, impedendo i danni e promuovendo l'equità, l'inclusione e la trasparenza (punti 2 e 28).

Il primo presupposto di una IA etica è che rispetti il valore giuridico della dignità umana, al quale fa riferimento l'articolo 2 del Trattato dell'UE e la Carta dei diritti fondamentali dell'UE (art. 1) e nel nostro ordinamento gli articoli 3 e 41 della nostra Costituzione.

La dignità è espressione dell'essenza e della natura dell'essere umano in quanto tale, le quali devono essere rispettate in qualunque contesto e attività con la conseguenza che, anche un gioco che «risvegli o rafforzi nel giocatore un'attitudine che neghi il diritto fondamentale di ogni persona ad essere riconosciuta e rispettata» [4], può costituire una violazione della dignità umana.

---

[3] Secondo il Gruppo indipendente di esperti di alto livello sull'Intelligenza Artificiale (istituito dalla Commissione Europea nel giugno 2018), *Orientamenti etici per una IA affidabile*, una IA affidabile si basa su tre componenti: legalità, eticità e robustezza.

[4] Causa C-36/02, *Omega Spielhallen- und Automatenaufstellungs-GmbH c. Oberbürgermeisterin*

Essa, in quanto base di tutti i diritti presenta «una connotazione oggettiva ed una dimensione estesa»[5].

Nella formula oggettiva la dignità umana impone che il suo rispetto prescinda anche dalla valutazione (soggettiva) che di essa ne faccia la persona stessa che ne è titolare.

In questo senso, emblematico il caso relativo al divieto emesso da un'autorità francese rispetto alla pratica del «lancio del nano»[6], attività alla quale si sottoponeva il ricorrente per guadagnarsi da vivere.

Sulla questione, il Comitato per i diritti umani delle Nazioni Unite, chiamato a pronunciarsi sulla legittimità dell'atto concludeva nel senso che lo Stato parte del Patto delle Nazioni Unite sui diritti civili e politici avrebbe dimostrato che «tutelare l'ordine pubblico, fa intervenire considerazioni sul piano della dignità umana»[7].

Sulla base di considerazioni superiori derivanti dal rispetto dovuto alla dignità della persona umana, si ritiene che possa vietarsi una certa attività, nonostante l'adozione di misure di protezione per garantire la sicurezza della persona interessata, e malgrado quest'ultima vi si presti volontariamente, per ottenerne una ricompensa.

Deriva che non sarebbe ammissibile un'attività che possa determinare una mortificazione della persona umana, neppure con il consenso del soggetto coinvolto. Invero, se il principio di dignità esalta l'autodeterminazio-

---

der Bundesstadt Bonn. Nel caso, per garantire la tutela dei valori fondamentali riconosciuti dall'articolo 1 della Costituzione tedesca, le autorità di Bonn avevano precluso ad una società tedesca la gestione di un "laserdromo" destinato alla pratica del «laser-sport», per il quale ci si avvaleva di attrezzatura fornita da una società britannica. La Corte di giustizia dell'UE nella sentenza Omega ammette che vi sia lesione della dignità umana tramite il «gioco di morte» e, per questo, giustifica la deroga alla libertà della prestazione di servizi.

[5] F.D. Busnelli, *Problemi giuridici di fine vita tra natura e artificio*, in *Riv. dir. civ.*, 2011, p. 161 ss.

[6] CE, Ass., 27 Octobre 1995, p. 372, *Case Commune de Morsang-sur-Orge*.

[7] CCPR/C/75/D/854/1999, 26 luglio 2002, Comunicazione n. 854/1999, *documents-dds-ny.un.org/doc/UNDOC/DER/G02/450/06/PDF/G0245006.pdf*.

ne, dall'altra parte la limita per escludere che ognuno possa trasformarsi in un potenziale tiranno di sé stesso[8].

Pertanto, qualsiasi sistema di IA che determini una qualche forma di avvilimento e sottomissione della persona alla tecnologia non dovrebbe ammettersi dovendo, in questo senso, escludersi anche solamente il rischio di "deumanizzazione", compresa quella resa attraverso processi decisionali interamente automatizzati[9].

La dignità quale caratteristica ontologica dell'essere umano impone che il suo valore prevalga rispetto agli algoritmi e ai dati su cui sono costruiti i sistemi di IA, consumando proprio intorno al merito della dignità il passaggio dalla supremazia dell'algoritmo (algocrazia) ad una dimensione morale di quest'ultimo.

In questa direzione, la predominanza dell'individuo sulle macchine deve mantenersi anche quando si considerano gli effetti che la rivoluzione digitale dispiega nel contesto lavorativo[10], dove la dignità, insieme all'autonomia del lavoratore, alla sua privacy e al suo diritto ad una giusta retribuzione, deve porsi come limite al potere tecnologico in favore di soluzioni il più inclusive possibili[11].

3. — *Autodeterminazione e protezione dei dati personali.*

Una delle componenti etico-giuridiche alle quali abbiamo sopra accennato è quella relativa al rispetto del valore dell'autonomia.

---

[8] V. Scalisi, *L'ermeneutica della dignità*, Milano, 2018.

[9] G. Noto La Diega, *Against the de-humanisation of Decision-Making*, in *JIPITEC*, 2018, 9 (3).

[10] Come affermato dall'European Group on ethices in Science and New technologies, *Future of work, Future of society*, 19 dicembre 2018, «together, technology, demographics and globalisation are amongst the most significant forces transforming the nature and role of work».

[11] G. Colaiacomo, *Intelligenza artificiale e dignità del lavoratore*, in D. Buzzelli, M. Palazzo (a cura di), *Intelligenza artificiale e diritti della persona*, 2022, pp. 205-222.

In termini generali, quando una persona deve essere coinvolta in una certa attività deve esserle richiesto, in applicazione al principio di autodeterminazione, un consenso libero e informato, quale proiezione del diritto alla libertà individuale.

L'istituto del consenso informato è previsto in molte fonti europee come l'art. 3, par. 2 della Carta dell'Unione Europea e in materia di protezione dei dati personali nell'articolo 8 della Carta dei diritti fondamentali dell'Unione Europea oltre che nel reg. UE 2016/679 (GDPR). Inoltre, nella Convenzione di Oviedo per l'attività biomedica (articolo 5); mentre all'interno del sistema giuridico nazionale viene collegato agli artt. 2, 3 e 32 della Costituzione.

Il principio del rispetto della volontà individuale è criterio giuridico che pervade tutte le ipotesi in cui una persona è chiamata ad esercitare una scelta, sia essa relativa al settore della salute o a una pratica commerciale, al trattamento dei dati personali o per la partecipazione all'attività di ricerca scientifica, e anche nel contesto di uso dei sistemi di IA.

In questo senso, nella Proposta di Regolamento europeo sull'IA è prescritto all'art. 52, par. 1, che «i fornitori garantiscono che i sistemi di IA destinati a interagire con le persone fisiche siano progettati e sviluppati in modo tale che le persone fisiche siano informate del fatto di stare interagendo con un sistema di IA, a meno che ciò non risulti evidente dalle circostanze e dal contesto di utilizzo»; fermo restando che, in alcuni casi, il consenso preventivo non sarà obbligatorio, come quando i sistemi di IA sono autorizzati dalla legge per accertare, prevenire, indagare e perseguire reati o, più in generale, quando sia necessario proteggere un interesse collettivo.

Lo stesso principio di autodeterminazione riguarda il tema del consenso alla raccolta e trattamento dei dati tramite tecnologie digitali.

Il rapporto tra IA e protezione dei dati personali è molto stretto essendo l'IA *data-driven*, cioè costruita intorno a dati e informazioni, sia inoltre perché sono numerose le applicazioni d'intelligenza artificiale rivolte ai trattamenti di dati personali.

Considerando che la protezione dei dati personali è uno dei diritti fondamentali, una analisi sulle riflessioni etiche dell'uso dell'IA non può prescindere dal considerare anche l'etica dei dati, vale a dire, come i dati vengono

creati, raccolti, curati, aggregati e mantenuti: il rispetto della legge e l'agire in modo appropriato e consono al proprio ruolo, adempiendo agli obblighi prescritti è una questione di etica dei dati[12].

Per la conformità etica e giuridica è necessario che il flusso dei dati personali volti ad alimentare l'IA risulti in linea con la normativa europea del GDPR il quale, all'art. 6, par. 1, lett. a), stabilisce che il trattamento è lecito solo se e nella misura in cui ricorra la condizione che l'interessato abbia espresso il consenso al trattamento dei propri dati personali per una o più specifiche finalità.

Deriva che il principio di autodeterminazione che si esplica nel diritto del singolo a decidere se ed entro quali limiti rendere i propri dati personali, resta valido anche nelle applicazioni di IA; tuttavia, alcune osservazioni si impongono dal momento che le caratteristiche strutturali dell'IA possono intaccare il valore della volontà individuale.

La prima è che anche quando il trattamento dei dati è giustificato dal consenso, dinanzi all'operare di un sistema di intelligenza ad autoapprendimento (machine learning), l'elaborazione che la macchina in via autonoma sceglie di compiere può cambiare in assenza di una qualsiasi supervisione umana le finalità del trattamento rispetto a quelle originarie per le quali il soggetto aveva manifestato la volontà, di fatto incidendo sull'attualità della base giuridica consensuale ed escludendo, in conclusione, la liceità del trattamento.

Ma se le finalità del trattamento dei dati si definiscono nel corso dell'attività stessa dal momento che l'IA apprende via via nuove informazioni e si evolve in base a queste, allora si impone una seconda riflessione.

Le finalità del trattamento cui sono destinati i dati personali nonché la base giuridica devono, come previsto dal GDPR, essere comunicati all'interessato, costituendo elementi dell'informativa (artt. 13 e 14)[13].

---

[12] P. Shaw, *The Law, Ethics and AI*, in M. Hervey, M. Lavy, *The Law of Artificial Intelligence*, 2021, pp. 31-66.

[13] Sulla valutazione della necessità dell'informativa agli utenti e di una idonea base giuridica per il trattamento dei dati, di recente il Garante Italiano per la protezione dei dati personali ha disposto (provvedimento del 30 marzo 2023, Registro dei provvedimenti n.

Invero, qualunque consenso presuppone, per essere volontario, un'attività di informazione. Il Regolamento in materia di protezione dei dati personali impone che le informazioni relative al trattamento siano fornite all'interessato in modo tale da essere facilmente accessibili e che sia utilizzato un linguaggio semplice e chiaro (art. 12) essendo lecito solo il trattamento che sia trasparente (art. 5, par. 1, lett. a)).

Deriva che la comprensione dovrebbe riguardare in generale il sistema di IA. Le parti interessate dovrebbero essere informate sulle loro interazioni con i sistemi di IA, e dovrebbe essere fatto in modo di consentire a coloro che sono interessati da un sistema di IA di comprenderne l'esito[14].

Tuttavia, la trasparenza non sempre è facile o possibile quando si usano strumenti di intelligenza artificiale.

La complessità dei sistemi di IA e il segreto della logica del processo decisionale determina la difficoltà per il titolare del trattamento dati di informare in modo intelligibile e facilmente accessibile il soggetto interessato con la conseguenza che «il diritto all'autodeterminazione informativa (di quest'ultimo) non potrebbe mai esser realmente effettivo»[15].

In conclusione, il potere di autodeterminazione si esprimerebbe nel diritto di ciascuno di scegliere sull'uso dei sistemi di IA e nella libertà di non avvalersi di essi, di poter permettere l'accesso alle proprie informazioni personali e di esercitare tutti i diritti di cui il proprietario dei dati deve rimanere padrone anche negli sviluppi della tecnologia dell'IA.

---

112) [doc. web n. 9870832 reperibile presso *www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870832*] la sospensione dell'uso di un sistema di IA. Nello specifico, il Garante ha adottato ai sensi dell'art. 58, par. 2, lett. f) del GDPR, nei confronti della società statunitense OpenAI L.L.C., quale titolare del trattamento dei dati personali effettuato attraverso il servizio ChatGPT di cui la prima è sviluppatrice e gestrice, la misura della limitazione provvisoria del trattamento dei dati personali degli interessati stabiliti nel territorio italiano, sul rilievo della carenza di informativa agli utenti e per l'assenza di idonea base giuridica in relazione alla raccolta dei dati personali e al loro trattamento per scopo di addestramento degli algoritmi sottesi al funzionamento di ChatGPT.

[14] Raccomandazione sull'Intelligenza Artificiale adottata dal Consiglio dell'OCSE riunito a livello ministeriale il 22 maggio 2019

[15] C. De Menech, *Intelligenza artificiale e autodeterminazione in materia sanitaria*, in *BioLaw Journal-Rivista di BioDiritto*, 2022, 1, pp. 181-203.

Il fatto che sia utilizzata l'IA deve essere indicato, così come che ci si avvalga di quei sistemi in un'operazione di trattamento dei dati, anche in considerazione dei rischi potenziali.

4. — *Principio di proporzionalità.*

I rischi che l'attività tecno-scientifica, compresa quindi l'IA e le sue evoluzioni, può comportare devono essere *proporzionati* e ragionevoli per i soggetti coinvolti.

Pertanto, un ulteriore principio etico-giuridico che deve essere rispettato è quello della proporzionalità.

In termini generali, il principio di proporzionalità impone un equilibrio tra mezzi adoperati e raggiungimento di un determinato fine. Ciò significa che deve essere rispettato un valore di misura per cui i primi non risultino eccessivi rispetto allo scopo da raggiungere[16].

Emblematico al riguardo l'articolo 16 della Convenzione di Oviedo che in materia di tutela delle persone che si prestano ad una ricerca stabilisce quale limite che i rischi che può correre la persona non siano *sproporzionati* in rapporto con i benefici potenziali della ricerca.

Allo stesso modo, gli strumenti di IA devono evitare rischi inutili e non possono ledere gli interessi fondamentali tutelati dall'ordinamento giuridico.

In questo senso, nella Proposta di Regolamento è prescritto che obblighi *proporzionati* possono essere imposti a tutti i partecipanti alla catena del valore per affrontare le varie fonti di rischio e assicurare un elevato livello di protezione dei diritti. Inoltre, nel momento in cui si diffonde e si utilizza una tecnologia di IA le restrizioni a certe libertà fondamentali devono essere proporzionate e limitate al minimo necessario, per prevenire e attenuare rischi gravi per la sicurezza e probabili violazioni dei diritti fondamentali.

---

[16] A. Langella, *Il Principio di Proporzionalità nella ricerca biomedica*, in questa *Rivista*, 2019, pp. 137-172.

Infine, nel Libro Bianco un richiamo al limite della proporzione è fissato prevedendo la possibilità di utilizzare sistemi di IA nella lotta alla criminalità nei limiti di un uso necessario e *proporzionato* (v. Conclusioni del Libro Bianco).

5. — *Un approccio basato sul rischio per l'IA. Prevenzione e precauzione.*

Nelle fonti dell'UE si chiarisce che l'uso dell'intelligenza artificiale può comportare rischi «che possono essere legati alle minacce informatiche, alla sicurezza personale (ad esempio in relazione ai nuovi usi dell'IA, come nel caso degli elettrodomestici), alla perdita di connettività, etc., e possono esistere al momento della commercializzazione dei prodotti o insorgere a seguito di aggiornamenti del software e dell'apprendimento automatico del prodotto quando questo è in uso» (cfr. Libro bianco sull'IA, par. 5).

I rischi che l'uso dell'intelligenza artificiale può provocare non sono considerati in modo esplicito dalla normativa.

Ciò dipende in primo luogo dal fatto che non è possibile determinare il pericolo con sufficiente certezza e quindi resta difficile prevederne una regolamentazione.

In termini generali, quando da un'attività possono derivare effetti potenzialmente insidiosi rispetto ai quali resta un grado di incertezza scientifica del verificarsi del danno, si deve tenere un comportamento precauzionale. E questo approccio è tanto più importante quanto più la pratica è in grado di mettere a repentaglio i diritti e le libertà fondamentali degli individui.

Quando invece si ha la certezza che possa verificarsi un danno, il principio di prevenzione, disciplinato per l'ambito del diritto dell'ambiente all'art. 192 TFUE, prevede la necessità di impedire l'evento dannoso.

A ciò, nei contesti di IA si provvede attraverso la predisposizione di ambienti sicuri e protetti, tecnicamente robusti per garantire che i sistemi, e gli ambienti in cui operano, non siano esposti ad usi malevoli, prestando particolare attenzione anche ai soggetti e ai gruppi vulnerabili (par. 51 degli Orientamenti etici per un IA affidabile, 2018).

Il principio di precauzione, che pure nasce nel contesto della materia della protezione dell'ambiente, e il cui primo riconoscimento a livello internazionale risale alla Carta mondiale della natura adottata dall'assemblea generale delle Nazioni Unite nel 1982, presuppone quali suoi requisiti l'identificazione di effetti potenzialmente negativi derivanti da un fenomeno, da un prodotto o da un procedimento e la valutazione scientifica del rischio che, per l'insufficienza dei dati, il loro carattere non concludente o la loro imprecisione, non consenta di determinare con sufficiente certezza il rischio in questione (Comunicazione della Commissione sul principio di precauzione COM (2000) 1 DEF).

A fronte di una incertezza del rischio è necessario che «i legislatori tengano in debita considerazione il principio di precauzione nella regolamentazione dell'IA» (Risoluzione del Parlamento europeo del 3 maggio 2022 sull'intelligenza artificiale in un'era digitale (2020/2266(INI)).

Gli operatori devono adottare delle misure precauzionali le quali devono essere proporzionate rispetto al livello di protezione desiderato a fronte della gravità della minaccia.

In questa direzione, la Proposta di regolamento, che adopera un approccio basato sul rischio, stabilisce requisiti diversi a seconda della natura del pericolo legato all'uso dell'IA. Prendendo in considerazione i rischi potenziali che possono derivare dall'uso dell'IA prescrive il divieto delle pratiche che possono risultare particolarmente dannose in quanto in contrasto con i valori dell'Unione (elenco di cui al titolo II); per sistemi "non ad alto rischio" impone obblighi di trasparenza limitati mentre per quelli "ad alto rischio", quelli cioè che pongono rischi significativi per la salute e la sicurezza o per i diritti fondamentali (cfr. Considerando 32 della Proposta di regolamento), prescrizioni più pesanti: documentazione e tracciabilità, trasparenza, sorveglianza umana, precisione e robustezza ed elevata qualità dei dati utilizzati (cfr. Considerando 43).

Entro il rispetto di questi obblighi, conformemente al criterio di proporzionalità, possono essere utilizzati nell'Unione solo i sistemi di IA ad alto rischio che non presentino rischi *inaccettabili* per gli interessi pubblici importanti dell'Unione (Considerando 27). Fermo restando che, il livello di

rischio "accettabile" per la società non può mai portare a ledere la dignità e l'integrità delle persone.

Nello specifico, sono considerati ad alto rischio i sistemi di IA relativi all'identificazione biometrica remota "in tempo reale" e "a posteriori" (cfr. Cons. 33); quelli relativi alla gestione e al funzionamento delle infrastrutture critiche (Cons. 34); all'istruzione e alla formazione professionale (Cons. 35); all'occupazione, alla gestione dei lavoratori e all'accesso al lavoro autonomo (Cons. 36). Inoltre, i sistemi di IA utilizzati per valutare il merito di credito o l'affidabilità creditizia delle persone fisiche (Cons. 37) e quelli utilizzati nella gestione della migrazione, dell'asilo e del controllo delle frontiere (Cons. 39).

Si configurano come particolarmente rischiosi anche i sistemi di IA destinati all'amministrazione della giustizia e ai processi democratici, in considerazione del loro impatto sulla democrazia, sullo Stato di diritto, sulle libertà individuali e sul diritto a un ricorso effettivo e a un giudice imparziale (Cons. 40).

In questo ultimo senso, l'uso dell'IA in materia giudiziaria potrebbe comportare numerosi vantaggi, primo fra tutti quello di ridurre il carico pendente presso i tribunali con conseguente accelerazione della risposta di giustizia ai consociati, tematica alla quale il nostro paese è particolarmente sensibile. Inoltre, gli algoritmi possono essere di supporto ad un'operazione di stabilizzazione della giurisprudenza e per rendere più controllabili i meccanismi decisori seguiti dal giudice[17], oltre che come mezzo attraverso il quale evitare errori e pregiudizi[18].

Tuttavia, numerosi casi hanno dimostrato che l'applicazione dell'IA al processo decisionale nel settore della giustizia può portare al rischio di giudizi non equi e discriminatori.

---

[17] Ad esempio, un algoritmo di machine-learning sviluppato dall'University College of London è stato in grado di prevedere l'esito dei casi della Corte Europea dei diritti dell'uomo basandosi esclusivamente sul contenuto testuale, con un'accuratezza del 79%; N. Aletras, D. Tsarapatsanis, D. Preotiuc-Pietro, V. Lampos, *Predicting judicial decisions of the European Court of Human rights: A natural language processing perspective*, in *PeerJ computer Science*, 2016, 1, 2.

[18] C. Cavaceppi, *L'intelligenza artificiale applicata al diritto penale: criticità attuali e prosepttive future,* in *Intelligenza artificiale. Algoritmi giuridici. Ius condendum o "fantadiritto"?*, a cura di G. Taddei Elmi e A. Contaldo, 2020, pp. 97-136.

È quanto avvenuto, ad esempio, a fronte dell'applicazione di algoritmi predittivi negli Stati Uniti (COMPAS, Correctional offender management profiling for alternative sanctions) e nel Regno Unito (HART, Harm Assessment Risk Tool)[19] per calcolare la probabilità di recidiva dei reati[20].

Utilizzando indici come la disoccupazione, l'abuso di sostanze e la mancanza di un alloggio per calcolare l'inclinazione a compiere reati, l'uso dell'algoritmo ha attestato un maggior grado di rischio di recidiva in certi gruppi di popolazione incorrendo in un pregiudizio (COMPAS commette discriminazione raziale calcolando una maggiore probabilità di recidiva per la popolazione afroamericana).

Un altro caso è quello relativo al sistema "Syri"[21] utilizzato – malgrado pareri negativi fossero già stati rilasciati dall'Autorità olandese per la protezione dei dati personali e il Consiglio di Stato – dal governo dei Paesi Bassi dal 2014 fino al 2020 per stabilire il rischio di frode fiscale o altri abusi da parte dei beneficiari di sussidi sociali. Come rilevato, l'algoritmo discriminava ingiustificatamente i cittadini meno facoltosi in quanto basato sulla premessa che fossero più a rischio i beneficiari di sussidi che vivevano in quartieri ad alta densità di residenti, a basso reddito, migranti e minoranze.

I casi in commento dimostrano che l'uso della tecnologia nel settore della giustizia solleva una serie di problemi etici legati all' inaccessibilità al sistema

---

[19] Sviluppato nell'ambito di una collaborazione tra il Durham Constabulary e l'Università di Cambridge, rappresenta uno dei primi impieghi operativi di metodi algoritmici delle forze dell'ordine del Regno Unito. Creato come parte di un programma noto come Checkpoint che prevede un regime alternativo all'azione penale per un sottogruppo specifico di autori di reati. M. Oswald, J. Grace, S. Urwin, G. C. Barnes, *Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality*, in *Information & Communications Technology Law*, 2018, 27, 2, pp. 223-250.

[20] Supreme Court of Wisconsin, *State of Wisconsin v. Eric L. Loomis*, Case no. 2015AP157-CR, 5 April-13 July 2016.

[21] C-09-550982-HA ZA 18-388, Corte distrettuale dell'Aia, *NCJM e al. e FNV contro lo Stato dei Paesi Bassi*, 6 Marzo 2020.
Per una ricostruzione si veda A. Rachovitsa, N. Johann, *The Human Rights implications of the use of AI in the digital welfare state: lessons learned from the Dutch SyRI case*, in *Human Rights Law Review*, 2022, 22, pp. 1-15. M. van Bekkum, F. Z. Borgesius, *Digital welfare fraud detection and the Dutch SyRI judgment* in *European Journal of Social Security*, 2021, 23(4), pp. 323–340.

e al pericolo di risultati arbitrari che portano alla discriminazione a danno della persona giudicata[22].

Questi aspetti sono considerati nella *Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi* adottata dalla Commissione europea per l'efficienza della giustizia (CEPEJ) nel 2018 e fissati al suo interno come principi etici.

6. — *Trasparenza e decisione automatizzata. Responsabilità.*

Le persone hanno diritto a conoscere le ragioni alla base delle decisioni che li riguardano e anche quando viene utilizzata una procedura informatica non può essere superato questo principio di trasparenza (v. punto 8.2, Cons. Stato, 8 aprile 2019, n. 2270).

Tuttavia, la decisione amministrativa gestita mediante algoritmo potrebbe non essere trasparente.

Ciò è accaduto, ad esempio, in Italia, in relazione alla gestione delle procedure di mobilità degli insegnanti effettuata attraverso l'uso di un software che avrebbe disposto i trasferimenti senza tener conto delle preferenze espresse e pur in presenza di posti disponibili nelle province indicate dagli insegnanti.

Queste decisioni sono state oggetto di una serie di sentenze del Consiglio di Stato (cfr. sentenze nn. 2270/2019, 8472/2019, 881/2020 e 1206/2021).

Le conclusioni alle quali arriva il Consiglio di Stato coincidono nell'indicare che l'utilizzo di un algoritmo che conduca ad una decisione finale deve essere conoscibile in ossequio ad una declinazione rafforzata del principio di trasparenza, che implica anche quello della piena conoscibilità di una regola espressa in un linguaggio differente da quello giuridico.

Per questa ragione, diventa fondamentale conoscere la logica sottesa all'algoritmo e le sue regole di operatività.

---

[22] Sebbene si possa rintracciare un numero molto più elevato di argomenti contro il processo decisionale algoritmico, come sottolinea G. Noto La Diega, *Against the dehumanisation of Decision-Making*, loc. cit.

In questo senso, il Conseil Consitutional francese (Decisione n. 2018-765 DC del 12 giugno 2018[23]), ha stabilito che quando i principi di funzionamento di un algoritmo non possono essere comunicati senza violarne il segreto, nessuna decisione individuale dovrebbe essere presa sulla base esclusiva dell'algoritmo (cfr., in particolare, i punti 70, 71, 72 della decisione). La sentenza francese impone due osservazioni: la prima è che "trasparenza" non significa pubblicare il codice sorgente né renderlo soggetto a una licenza aperta, quanto piuttosto garantire che la logica impiegata dai tecnologi e la logica adottata dalla tecnologia siano chiare[24]. La seconda riguarda il *principio di non esclusività* della decisione automatizzata.

Ciò si ricava dall'art. 22 GDPR che riconosce il diritto dell'interessato di "non essere sottoposto a una decisione basata *unicamente* sul trattamento automatizzato, compresa la profilazione[25], che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona". Del resto, "vagliare, catalogare, valutare per punteggio, aggregare, condizionare o manipolare" le persone entra in collisione anche con il concetto di dignità umana (Orientamenti etici per un'IA affidabile, par. 41).

Nel GDPR, richiami al principio di trasparenza si rintracciano oltre che nel già citato art. 12 (v. sopra), al Considerando 39, il quale raccomanda di rendere trasparenti le "modalità con cui i dati sono raccolti, utilizzati, consultati o altrimenti trattati".

In materia di IA il riferimento va al Piano coordinato sull'intelligenza artificiale (par. 2.6 del Piano coordinato sull'intelligenza artificiale, Comunicazione della Commissione COM (2018) 795 final) che afferma «fondamentale che gli umani comprendano in che modo l'IA prende le decisioni» (par. 2.6,

---

[23] Reperibile in *www.conseil-constitutionnel.fr/en/decision/2018/2018765DC.htm*.

[24] P. Shaw, *The Law, Ethics and AI*, cit., p. 57.

[25] L'art. 4 (4) GDPR definisce «profilazione» qualsiasi forma di trattamento automatizzato di dati personali consistente nell'utilizzo di tali dati personali per valutare determinati aspetti personali relativi a una persona fisica, in particolare per analizzare o prevedere aspetti riguardanti il rendimento professionale, la situazione economica, la salute, le preferenze personali, gli interessi, l'affidabilità, il comportamento, l'ubicazione o gli spostamenti di detta persona fisica.

COM (2018) 795 final); essendo « importante registrare e documentare sia le decisioni prese dai sistemi sia l'intero processo (compresa una descrizione della raccolta e dell'etichettatura dei dati e una descrizione dell'algoritmo utilizzato) che ha portato alle decisioni» (Comunicazione del 2019 "Creare fiducia nell'intelligenza artificiale incentrata sull'uomo" (cfr. par. 2.2.IV)).

Di fatto, la trasparenza impone di adottare delle misure per rendere comprensibile il processo decisionale algoritmico. Così facendo, assume una dimensione etica quale presupposto per consentire all'uomo di "governare" l'IA, tutelare la dignità umana ed essere funzionale all'autodeterminazione.

D'altra parte, l'inintelligibilità del procedimento/risultato dell'IA renderebbe impossibile per l'operatore umano scegliere tra accettare la soluzione proposta dalla macchina o percorrere una diversa conclusione; dovendo anche considerarsi che questa seconda alternativa potrebbe essere "ostacolata" dalla fiducia pressoché illimitata che l'uomo nutrirebbe verso la macchina, sulla base di una presunta infallibilità di quella [26].

Nel settore sanitario, questa incomprensibilità può avere come effetto un inaridimento, un declino delle abilità pratiche del medico ed altri rischi inerenti alla relazione con il paziente: il medico non sarebbe in grado di fornirgli la spiegazione relativa la scelta terapeutica suggerita o la valutazione diagnostica effettuata dalla macchina [27]. In pratica, entrambi i soggetti reste-

---

[26] J.M. Logg, J.A. Minson, A. Moore, *Algorithm appreciation: people prefer algorithmic to human judgment*, in *Organizational behavior and human decision processes*, 2019, 151, pp. 90-103.

[27] Comitato nazionale per la biosicurezza, le biotecnologie e le scienze della vita, *Intelligenza Artificiale e medicina: aspetti etici*, 29 maggio 2020. Come avverte la dottrina, sono molti i rischi che l'etica medica ha iniziato a evidenziare: preoccupazioni riguardanti pregiudizi algoritmici, trasparenza, disumanizzazione delle relazioni medico-paziente. K. Astromskė, E. Peičius, P. Astromskis, *Ethical and legal challenges of informed consent applying artificial intelligence in medical diagnostic consultations*, in *AI & Soc*, 2021, 36, pp. 509-520; A. Froomkin, I. Kerr, J. Pineau, *When AIs outperform doctors: confronting the challenges of tort-induced over-reliance on machine learning*, in *Arizona Law Rev.*, 2019, 61(1), pp. 33-100; DS Char, N.H. Shah, D. Magnus, *Implementing machine learning in health care-addressing ethical challenges*, in *N Engl J Med*, 2018, 378(11), pp. 981-983. Si veda inoltre il Rapporto del Comitato direttivo sui diritti umani nel campo della biomedicina e della salute (CDBIO) del Consiglio d'Europa sull'impatto dell'IA sulla relazione medico-paziente, reperibile in inglese presso *rm.coe.int/inf-2022-5-report-impact-of-ai-on-doctor-patient-relations-e/1680a68859*. Il rapporto esamina il

rebbero sopraffatti dalla tecnica. Tuttavia, non sarebbero escluse eventuali responsabilità per il medico perché, coerentemente con *il principio di non esclusività*, le decisioni basate sui dati tecnico-scientifici devono essere attribuibili alle persone, le quali restano responsabili in caso di mancato rispetto delle norme e dei principi.

Del resto, nell'innovazione del lavoro mediante l'IA, la collaborazione antropo-meccanica non esclude il rispetto del principio *primum non nocere*, la violazione del quale evoca responsabilità. E alla responsabilità fa riferimento la Comunicazione "Costruire la fiducia nell'intelligenza artificiale incentrata sull'uomo" (cfr. par. 2.2.VII) secondo la quale "devono essere messi in atto meccanismi per garantire la responsabilità e l'affidabilità dei sistemi di IA e dei loro risultati, sia prima che dopo l'implementazione. La possibilità di effettuare audit dei sistemi di IA è essenziale, poiché la valutazione dei sistemi di IA da parte di revisori interni ed esterni e la disponibilità di rapporti di valutazione contribuiscono notevolmente all'affidabilità della tecnologia. La possibilità di audit esterni dovrebbe essere garantita soprattutto per le applicazioni che incidono sui diritti fondamentali, ad esempio quelle critiche per la sicurezza".

In materia di robotica, come detto, poter decifrare il modo di operare o di compiere azioni da parte di un robot, tanto da poterne anticipare i comportamenti, potrebbe significare non esporsi mai a eventuali situazioni dannose[28] (v. quanto detto sopra in materia di principi di prevenzione e precauzione).

In questo senso, il discorso torna sulla trasparenza, che è l'argomento dal quale siamo partiti, e rispetto al quale la risoluzione del Parlamento europeo del 2017 sulla robotica (2018/C 252/25) pone l'accento affermando che

---

potenziale impatto dell'IA sui diritti umani secondo sei temi: (i) disuguaglianza nell'accesso a cure sanitarie di alta qualità, (ii) trasparenza per gli operatori sanitari e i pazienti, (iii) rischio di pregiudizio sociale nei sistemi di IA, (iv) indebolimento della considerazione del benessere del paziente, (5) rischio di condizionamento dell'automazione, dequalificazione e spostamento della responsabilità e (v) impatto sul diritto alla privacy.

[28] R. Trezza, *Diritto e intelligenza artificiale. Etica – Privacy – Responsabilità – Decisione*, Pisa, 2020, p. 108.

«dovrebbe sempre essere possibile indicare la logica alla base di ogni decisione presa con l'ausilio dell'intelligenza artificiale che possa avere un impatto rilevante sulla vita di una o più persone […]; essere sempre possibile ridurre i calcoli del sistema di intelligenza artificiale a una forma comprensibile per gli esseri umani; e […] i robot avanzati dovrebbero essere dotati di una 'scatola nera' che registri i dati di tutte le operazioni eseguite dalla macchina, compresi, se del caso, i passaggi logici che hanno portato alla formulazione delle sue decisioni» (par. 12).

In conclusione, nella misura in cui si esige la trasparenza quale presupposto attraverso il quale l'uomo può effettuare un controllo sull'IA (cfr. Cons. 48 della Proposta di regolamento secondo cui "i sistemi di IA ad alto rischio dovrebbero essere progettati e sviluppati in modo da consentire alle persone fisiche di *sorvegliarne il funzionamento*") si ammette che la «tecnologia [deve essere] funzionale e servente all'uomo[29]» la cui essenza resta insostituibile e il cui contributo nei processi decisionali dovrebbero rimanere sempre protetto.

In fondo, come detto, nella dignità dell'uomo, che è il suo valore più alto, rientra anche la "dignità della decisione" che è il valore più alto dell'intelletto[30].

In questo senso, i casi presi in discussione circa l'uso dell'IA per le decisioni giudiziarie e nel contesto della pratica medica, dimostrerebbero che i sistemi della tecnologia potrebbero rivelarsi inefficienti per la mancanza di certe prerogative che sono sole dell'uomo e non della macchina: invero, pur essendo l'IA in campo medico un mezzo promettente per tagliare i costi, ridurre i tempi di attesa o riempire l'esistente lacuna nella copertura in cui l'accesso agli operatori sanitari e alle istituzioni è limitato, l'IA potrebbe fornire cure di qualità inferiore, ad esempio per quanto concerne le interazioni faccia a faccia (Rapporto cit., p. 33).

Allo stesso modo, nelle decisioni giudiziarie unicamente automatizzate potrebbe mancare quel «concetto umano di giustizia [il quale] non può es-

---

[29] R. Trezza, *Diritto e intelligenza artificiale. Etica – Privacy – Responsabilità - Decisione*, cit., p. 27.

[30] R. Trezza, *Diritto e intelligenza artificiale. Etica-privacy-responsabilità-decisione*, loc. cit.

sere separato dalle convinzioni personali, dalle esperienze e dalle emozioni dell'essere umano concreto»[31].

7. — *Equità e correttezza dei sistemi di IA*

L'art. 14 del d.lgs. 196/2003 (oggi abrogato) stabiliva che «nessun atto o provvedimento giudiziario o amministrativo che implichi una valutazione del comportamento umano può essere fondato unicamente su un trattamento automatizzato di dati personali volto a definire il profilo o la personalità dell'interessato».

Nel GDPR i processi decisionali automatizzati sono presi in considerazione nell'art. 22 che riconosce, come detto, il diritto dell'interessato di «non essere sottoposto a una decisione basata *unicamente* sul trattamento automatizzato, compresa la profilazione[32], che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona».

La profilazione è tecnica utilizzata in diversi settori (nelle pratiche di marketing, in ambito pubblico, della giustizia (v. i casi sopra)).

Anche nel fenomeno della *filter bubble*[33] che è effetto di un uso automatizzato dei dati attraverso algoritmi[34] solleva, in generale, problemi di

---

[31] T. Ćapeta, *Of Judges and Robots*, in M. Ilešič, *Challenges of Law in Life Reality*, 2017, pp. 129, 138.

[32] L'art. 4 (4) GDPR definisce «profilazione» qualsiasi forma di trattamento automatizzato di dati personali consistente nell'utilizzo di tali dati personali per valutare determinati aspetti personali relativi a una persona fisica, in particolare per analizzare o prevedere aspetti riguardanti il rendimento professionale, la situazione economica, la salute, le preferenze personali, gli interessi, l'affidabilità, il comportamento, l'ubicazione o gli spostamenti di detta persona fisica.

[33] E. Paiser, *The filter bubble. What internet is hiding from you*, Penguin Press, New York, 2011.

[34] Utili per capire di cosa si tratta le definizioni ripotate da M. Bianca, *La filter bubble e il problema dell'identità digitale*, in *Media Laws*, 2019, 2, p. 39 ss. Per l'Oxford Dictionary, è «una situazione in cui un utente di Internet incontra solo informazioni e opinioni che si

natura etica, *in primis*, strappi all'autonomia e violazioni al valore di identità personale[35], che è il diritto fondamentale che vengano rappresentati con i propri reali caratteri e senza travisamenti, le idee, lo stile di vita, il proprio patrimonio intellettuale, ideologico, etico, professionale[36].

Il principio di autodeterminazione viene calpestato nella misura in cui le procedure automatizzate di dati prescindono dal consenso e, inoltre, in quanto limitano di fatto la libertà di scelta proponendo un numero inferiore di offerte commerciali, dati, notizie cui l'utente potrebbe accedere senza il filtro della profilazione.

Consegue che tramite la profilazione si perpetua un'alterazione dell'identità personale dell'individuo: ogni volta che, a causa della profilazione, solo alcune informazioni vengono fornite dal sistema, il meccanismo perpetua una manipolazione dell'identità dell'utilizzatore.

Inoltre, occorrendo «valutare gli effetti dello sviluppo, della distribuzione e dell'utilizzo di un sistema di IA sugli individui e anche l'impatto sociale, tenendo conto degli effetti del sistema sulle istituzioni, sulla democrazia e sulla società in generale» (Orientamenti etici, par. 43 e 86), si dovrebbe censurare la profilazione per la deriva antidemocratica che rappresenta dal momento che incide negativamente sulla formazione della coscienza politica e collettiva[37].

Nell'ambito della profilazione e delle decisioni automatizzate assume centralità anche il rispetto del principio di non discriminazione.

Un'IA etica è un'IA equa, vale a dire una IA che non discrimina (art. 21 della Carta dei diritti fondamentali dell'UE e art. 3 Cost.).

Difatti occorre «garantire che gli individui e i gruppi siano liberi da di-

---

conformano e rafforzano le proprie convinzioni, causata da algoritmi che personalizzano l'esperienza online di un individuo».

Per Techopedia: «La *filter bubble* è l'isolamento intellettuale che può verificarsi quando i siti web utilizzano algoritmi per ipotizzare in modo selettivo le informazioni che un utente vorrebbe vedere, e poi forniscono informazioni all'utente in base a questa supposizione».

[35] M. Bianca, *La filter bubble*, loc. cit.

[36] A. Torrente, P. Schlesinger, *Manuale di diritto privato*, 24ma ed., Milano, 2019, p. 148.

[37] M. Bianca, *La filter bubble*, cit., p.45.

storsioni inique, discriminazioni e stigmatizzazioni» come sottolineato a livello europeo negli "Orientamenti etici per un'IA affidabile" (cfr. par. 52).

Ogni qualvolta il sistema sia fonte di tali alterazioni la dimensione procedurale dell'equità (cfr. ancora Orientamenti) «implica la capacità di impugnare le decisioni elaborate dall'IA».

Mentre, in termini sostanziali, l'interrogativo diventa se e come l'equità possa essere automatizzata, e quale definizione di equità debba essere applicata e trasferita nell'algoritmo.

La creazione di un'equità algoritmica risulterebbe impresa difficile dal momento che lo stesso concetto giuridico di ciò che è "equo" ovvero "discriminatorio" è a livello europeo flessibile e sempre "contestualizzato" nella giurisprudenza, con la conseguenza che sviluppatori di sistemi, controllori, regolatori e utenti non dispongono di requisiti legali chiari, stabili e coerenti che siano traducibili in meccanismi di progettazione e di governance del sistema per individuare, rimediare e prevenire la discriminazione automatica[38].

Tuttavia, il più grande limite sarebbe rappresentato dal fatto che rispetto al processo decisionale umano gli algoritmi non sono altrettanto *intuitivi* per automatizzare una equità rispettosa dello standard flessibile europeo[39].

Allora se, come scrisse Albert Einstein, la mente *intuitiva* è un dono sacro e la mente razionale è un fedele servo, alcuni timori che si paventano a fronte della diffusione massiva dei sistemi di IA potrebbero essere – almeno in parte – sopiti.

---

[38] S. Wachter, B. Mittelstadt, C. Russell, *Why fairness cannot be automated: bringing the gap between EU non-discrimination law and AI*, in *Computer Law & Security Review*, 2021, 41, *passim*.

[39] S. Wachter, B. Mittelstadt, C. Russell, *Why fairness cannot be automated: bringing the gap between EU non-discrimination law and AI*, cit., *passim*. Secondo gli autori la flessibilità dell'equità in termini giuridici ha lasciato aperta la questione di quali test debbano essere utilizzati per valutare la discriminazione nella pratica. Difronte al tentativo degli informatici di trovare metriche di equità adeguate, che siano conformi alla legge e al contempo abbastanza statici da poter essere codificati, le comunità legali e tecnica potrebbero trarre beneficio da un dialogo più stretto su come progettare considerazioni europee di uguaglianza ed equità nell'IA.