LUISS Guido Carli
LIBERA UNIVERSITÀ INTERNAZIONALE DEGLI STUDI SOCIALI

Dipartimento di *Giurisprudenza*

Cattedra di *Tutela dei diritti umani*

L'uso degli algoritmi nel processo penale e la sua conformità ai principi dell'equo processo

RELATORE

Prof. Pietro Pustorino

CANDIDATO

Elisa Galloppa

CORRELATORE

Prof. Maurizio Bellacosa

MATRICOLA

154273

Anno Accademico 2020-2021

INDICE

CAPITOLO I

INTELLIGENZA ARTIFICIALE: NOZIONI DI RIFERIMENTO

CAPITOLO II

IL QUADRO GIURIDICO DI RIFERIMENTO IN MATERIA DI INTELLIGENZA ARTIFICIALE

CAPITOLO III

ORDINAMENTO ITALIANO ED INTELLIGENZA ARTIFICIALE

CAPITOLO V

IMPATTO DEGLI ALGORITMI SUL DIRITTO ALL'EQUO PROCESSO

INTRODUZIONE

L'Intelligenza Artificiale pervade il nostro vivere quotidiano al punto da prefigurare una società "algoritmica" che sancisce la definitiva transizione nell'era digitale. Molteplici sono gli utilizzi delle applicazioni basate sull'intelligenza artificiale: è infatti un algoritmo a suggerire i prodotti che potremmo apprezzare maggiormente durante gli acquisti online, a permettere ai computer che usiamo di funzionare, a guidare i veicoli senza conducente e a suggerire il percorso più breve per giungere a destinazione, semplificando notevolmente le nostre attività quotidiane.

Le applicazioni non si limitano solo al mondo dell'*e-commerce*, dell'istruzione, dell'informatica, della medicina o dell'intrattenimento: a titolo esemplificativo, la piattaforma Netflix utilizza algoritmi per suggerire film in linea con quelli visti di recente, la piattaforma Spotify crea playlist sulla base delle canzoni che ascoltiamo più frequentemente. Numerose applicazioni degli algoritmi si possono rinvenire anche nel campo della finanza (per l'assegnazione del credito), delle risorse umane (per le assunzioni e il reclutamento di personale) e dell'istruzione. Da ultimo sono stati impiegati anche nel settore sanitario quali strumenti per la gestione dell'emergenza pandemica. Basti pensare che in Italia si è fatto ricorso ad un algoritmo per determinare i colori delle regioni in funzione dell'indice di contagio da Covid-19 ai fini dell'applicazione delle misure di contenimento.

Per comprendere il potenziale applicativo di questi modelli matematici e l'impatto che hanno sulla nostra vita (al punto che Eric Sadin nel libro "Critica della ragione artificiale. Una difesa dell'umanità" ha parlato di «*assistentato algoritmico del nostro quotidiano*»), basti pensare che recentemente è stato realizzato un algoritmo in grado di capire, tramite le espressioni del volto del lettore, se il soggetto in questione abbia realmente compreso il testo, ricevendo ed elaborando i segnali emessi dal cervello del lettore.

Se fino a qualche decennio fa un simile traguardo sarebbe risultato fantascientifico, nell'arco di un breve lasso temporale è divenuto realtà. Gli

algoritmi riescono ormai a comprendere atti e fatti di un essere umano che gli umani stessi non riuscirebbero a cogliere, arrivando a conoscerci più di quanto noi stessi riusciamo a fare e risolvendo in modo efficiente problemi estremamente complessi anche per le nostre capacità.

Nel campo del diritto, che è l'ambito esaminato nella tesi, i sistemi di intelligenza artificiale pongono nuovi e inaspettati quesiti. Ci troviamo di fronte a nuove entità, le cui decisioni e risultati non sono integralmente l'effetto di azioni umane. Esse derivano da una serie di processi che, a seguito dell'istruzione della macchina, assumono autonoma capacità decisionale.

Scopo del presente lavoro è quello di fornire al lettore strumenti di analisi dell'attuale scenario applicativo degli algoritmi, con particolare riguardo al loro impiego in ambito processuale. Il *fil rouge* che lega la trattazione è il tema dell'ambivalenza degli strumenti di valutazione del rischio di recidiva dell'imputato, la cui applicazione si muove su un sottile crinale: quello tra mero supporto alle valutazioni del giudice e strumento delegato a decidere sulla libertà o reclusione del soggetto.

Nel Capitolo I si forniranno le definizioni utili alla comprensione della materia trattata, con particolare riferimento al concetto di "algoritmo" e al meccanismo del "*machine learning*", la principale tecnica di apprendimento degli algoritmi.

Il Capitolo II sarà dedicato all'analisi del quadro giuridico internazionale ed europeo in materia di intelligenza artificiale, con uno specifico *focus* sui principi contenuti nel Regolamento generale sulla protezione dei dati e nella Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi.

Nel Capitolo III si prospetteranno i recenti arresti della giurisprudenza in tema di uso degli algoritmi nel processo amministrativo. Nel prosieguo si analizzeranno le preclusioni a livello costituzionale e penale dell'ordinamento nazionale, ostative all'ingresso degli algoritmi nel processo penale italiano.

Il Capitolo IV fornirà un quadro dell'attuale impiego degli applicativi di intelligenza artificiale in ambito giudiziario, con riferimento alla diffusione degli strumenti di polizia predittiva in Europa. Finalità dell'analisi svolta è dimostrare

che il fenomeno degli algoritmi predittivi non è uno scenario *in itinere*, ma è già una realtà anche nel contesto europeo. I *software* di *predictive policing* costituiscono una delle possibili declinazioni degli strumenti di previsione del rischio. Pur essendo relegati alla fase investigativa, è vitale mantenere vivo il dibattito e l'attenzione sull'attuale stato dell'arte, per evitare di subire acriticamente le conseguenze del progresso in ambiti sensibili, che rischiano di minare i diritti umani.

La seconda parte del capitolo sarà dedicata all'analisi del sistema statunitense, quello che più di ogni altro ha accolto con entusiasmo l'utilizzo degli algoritmi predittivi nel processo penale. Verrà presentato il celebre caso *State vs Loomis*, emblematico dell'impiego lesivo dei *risk assessment tool* in fase di comminazione della pena. Nel caso in questione, la Corte Suprema del Wisconsin ha sostanzialmente ammesso la legittimità degli strumenti predittivi in giudizio, a patto che non costituiscano l'unico elemento su cui si fonda la decisione. Come si cercherà di dimostrare, tale pronuncia costituisce un precedente rischioso in tal senso: il risultato dell'algoritmo può non essere l'elemento decisivo per la condanna dell'imputato, ma è sicuramente quello che influenza maggiormente le determinazioni del giudice in senso sfavorevole all'accusato.

Su tali premesse si apre il Capitolo V. Dopo aver delineato le criticità sottese all'uso dei *risk assessment tool* in fase cautelare e di *sentencing,* la trattazione proseguirà con l'analisi dell'impatto degli algoritmi sul diritto all'equo processo contenuto nell'art. 6 della Convenzione europea per la salvaguardia dei diritti dell'uomo e delle libertà fondamentali. Richiamando la giurisprudenza della Corte EDU, si tenterà di dimostrare i numerosi profili di incompatibilità tra le garanzie a fondamento del giusto processo e gli strumenti di valutazione del rischio.

Infine, si evidenzierà come gli algoritmi, considerati strumenti neutrali e oggettivi in virtù di un *automation bias*, nascondano in realtà numerosi pregiudizi razziali: l'effetto è deleterio rispetto al divieto di discriminazione e alla presunzione di innocenza.

In definitiva, l'elaborato si pone in prospettiva critica, ma non preclusiva, rispetto alle recenti aperture del mondo giudiziario agli strumenti di intelligenza artificiale.

L'obiettivo primario è richiamare l'attenzione del lettore sulle surrettizie violazioni che tali strumenti possono determinare per i diritti fondamentali.

CAPITOLO I

INTELLIGENZA ARTIFICIALE: NOZIONI DI RIFERIMENTO

1.1 Il concetto di Intelligenza Artificiale

Il successo evolutivo della specie umana (classificata non a caso con il termine *homo sapiens*, propr. «uomo sapiente») è stato attribuito alle sue straordinarie capacità cerebrali, che le hanno fornito le abilità per primeggiare sulle altre specie, *in primis* tra tutte la capacità di linguaggio, di cooperazione, di elaborare un pensiero complesso, di coscienza e di astrazione.

Nonostante sia un tratto così caratterizzante e connaturato al nostro essere, il concetto di *intelligenza* è tutt'altro che intuitivo proprio in virtù della complessità dei fattori coinvolti, tra cui la capacità di elaborare pensieri, giudizi e soluzioni, la capacità di memoria, di comprensione e di apprendimento. Pertanto risulta particolarmente complesso individuare una definizione univoca del summenzionato concetto, anche in virtù del fatto che esistono diverse tipologie di intelligenza[1]: lo psicologo statunitense Howard Gardner aveva teorizzato nel 1983 la *Teoria delle intelligenze multiple*, secondo la quale se ne potevano addirittura individuare sette tipologie (nello specifico: l'intelligenza spaziale, l'intelligenza sociale, l'intelligenza introspettiva, l'intelligenza corporeo -cinestetica, l'intelligenza musicale).[2]

Così come il concetto di intelligenza umana, anche quello di intelligenza artificiale ha posto altrettante, se non più complesse, sfide di definizione.

[1] S. Quintarelli, *Intelligenza Artificiale: cos'è davvero, come funziona, che effetti avrà*, Bollati Boringhieri, Torino, 2020, p. 33
[2] H. Gardner, *Frames of mind: The theory of multiple intelligences,* Basic Books, New York, 2011

L'espressione fu utilizzata per la prima volta nel 1956 da John McCarthy[3] in occasione di un convegno al Dartmouth College di Hannover[4] , descrivendolo come un processo consistente nel far sì che "una macchina si comporti in modi che sarebbero definiti intelligenti se fosse un essere umano a comportarsi così".

Nel tempo sono state fornite numerose definizioni del concetto e , pur non potendo pervenire a una definizione universale, in estrema sintesi possiamo definirla come l'insieme dei metodi scientifici, di teorie e di tecnologie il cui obiettivo è quello di riprodurre ,attraverso una macchina, le abilità cognitive degli esseri umani.[5] La definizione sottende una concezione "antropomorfica" della macchina, che sarebbe così in grado di riprodurre determinate funzioni cerebrali appannaggio di un essere vivente dotato di logica.

Il primo a mostrare un approccio funzionale di studio al problema dell'intelligenza di una macchina e dei suoi limiti fu il matematico Alan Turing[6] che nel 1950 illustrò nell' articolo *Computing Machinery and Intelligence,* apparso sulla rivista *Mind*, il celebre *Test di Turing*[7] in cui si domandava se le macchine fossero in grado di pensare. La macchina avrebbe passato il test se l'esaminatore umano, dopo avergli posto alcune domande, non fosse stato in grado di distinguere se le risposte provenissero da una macchina o da un essere umano.[8] Alla luce del test di Turing la macchina deve possedere determinate capacità per essere considerata *intelligente*: capacità di interpretazione del linguaggio naturale (per comunicare con l'esaminatore), capacità di rappresentazione della conoscenza (per memorizzare ciò che sa), di ragionamento automatico (per utilizzare la conoscenza

[3] Informatico statunitense che ha vinto il Premio Turing nel 1971 per i suoi contributi nel campo dell'intelligenza artificiale.
[4] Nello specifico, oggetto del convegno era uno studio circa la possibilità di sviluppare macchine "intelligenti", in grado di utilizzare il linguaggio, formulare astrazioni, risolvere i problemi complessi e migliorare le proprie prestazioni nel tempo.
[5] CEPEJ*, Carta etica sull'uso dell'intelligenza artificiale nei sistemi giudiziari e nel loro ambiente*, 2018
[6] Matematico e crittografo nato a Londra nel 1912 e considerato uno dei padri fondatori dell'informatica e dell'intelligenza artificiale, celebre soprattutto per avere decifrato i codici utilizzati dai tedeschi durante la seconda guerra mondiale tramite il sistema Enigma
[7] Noto anche come *The imitation game* (il gioco dell'imitazione)
[8] S. J. Russell, P. Norvig, A*rtificial Intelligence. A Modern Approach*, Third Edition, Prentice Hall, 2010, p. 5-6

memorizzata e trarre nuove conclusioni) e di apprendimento (per individuare e estrapolare dei pattern)[9]. A queste quattro caratteristiche fondamentali se ne possono aggiungerne due: la visione artificiale per percepire gli oggetti nell'ambiente circostante e la robotica per manipolarli o spostarli fisicamente.

Dopo numerosi tentativi, il test di Turing è stato superato nel 2016 da un algoritmo che è stato in grado di riprodurre all'interno di un video dei suoni, risultati indistinguibili per gli spettatori rispetto a quelli prodotti da un essere umano, dopo che gli studiosi gli avevano sottoposto circa 46 .000 diversi sound che l'algoritmo era stato in grado di apprendere e ricreare dai diversi frammenti all'interno di un suono coesivo[10].

È possibile classificare l'intelligenza artificiale in "ristretta" e "generale". Per intelligenza artificiale "ristretta" si intendono macchine che agiscono come se fossero intelligenti, cioè in grado di eseguire compiti specifici seppur complessi, come giocare a scacchi o guidare un veicolo.[11] La macchina è dotata di numerose capacità: pianificazione del movimento, elaborazione del linguaggio naturale, riconoscimento di un discorso, visione artificiale[12].

L'intelligenza artificiale "forte" implica invece un passaggio ulteriore: queste macchine sono dotate di creatività, intuito, emozioni, capacità di astrazione e coscienza[13]. Sono capaci di contestualizzare problemi specializzati di varia natura in maniera completamente autonoma[14], adattandosi a risolvere qualsiasi compito gli venga assegnato a prescindere dal contesto di inserimento[15] .Tale concezione postula la possibilità di realizzare macchine che in futuro potrebbero essere in grado di superare le capacità umane: come afferma Sarle[16] "*il calcolatore non è semplicemente uno strumento per lo studio della mente, ma piuttosto, quando sia*

[9] *Ibidem*
[10] Malik, Iman, Ek Carl Henrik, *Neural Translation of Musical Style,* arXiv, 2017.
[11] S. Quintarelli, *ult. op. cit.,* p. 36
[12] Berryhill, Jamie, Kévin Kok Heang, Rob Clogher and K. McBride. "*Hello, World: Artificial intelligence and its use in the public sector.*" (2019), p.5
[13] *Supra* nota 12, pp. 9-10
[14] CEPEJ*, Carta etica sull'uso dell'intelligenza artificiale nei sistemi giudiziari e nel loro ambiente*, 2018
[15] *Ibidem*
[16] Professore di filosofia all'Università di Berkley

programmato opportunamente, `e una vera mente". Sebbene questo tipo di intelligenza sia ancora una utopia e la totalità dell'intelligenza artificiale di cui disponiamo rientri nella prima classificazione, in un futuro prossimo potrebbe divenire una realtà.

Numerose criticità si sono riscontrate circa la possibilità di paragonare l'intelligenza artificiale a quella umana: di fatto, pur essendo i padri di queste macchine, il pensiero di vedere replicate funzioni che sentiamo legate alla nostra più intima essenza di esseri umani provoca ancora un senso di disagio[17].

Il filosofo francese Eric Sàdin nel suo libro *Critica alla ragione artificiale* sostiene che non sia corretto modellare il principio di una intelligenza computazionale sulla nostra, poiché tra l'una e l'altra non può esistere un rapporto di similitudine per due ragioni:[18] la prima che è l'intelligenza artificiale è priva di corpo e si limita a ridurre a un codice binario determinati elementi del reale; la seconda è che l'intelligenza, per definizione, ha bisogno di una relazione aperta con gli esseri e le cose per potersi evolvere e distinguere. Per quanto riguarda il problema della coscienza nella macchina, Searle ha individuato il nodo centrale della questione affermando che la simulazione di un processo cognitivo non può produrre gli stessi effetti della neurobiologia di quel processo cognitivo. Di conseguenza, se anche la macchina dovesse dar prova di una qualche forma di coscienza o intenzionalità, non sarebbe il prodotto di un processo assimilabile al nostro ma una mera simulazione della nostra coscienza[19] .

Pur non potendo fornire una soluzione definitiva alla domanda "può una macchina pensare?", è innegabile che uno dei meriti principali dell'Intelligenza Artificiale sia proprio quello di essere una scienza sperimentale, basata sull'implementazione piuttosto che sulla mera simulazione di processi computazionali. La sfida non deve essere quella di duplicare la mente umana, resa

[17] A. Carcaterra , *Machinae autonome e decisione robotica* in *Decisione robotica,* A. Carleo (a cura di) Il Mulino , Bologna, 2019, p. 33
[18] E. Sadin, *Critica della ragione artificiale: una difesa dell'umanità*, Luiss University Press, Roma, 2019
[19] Searle, J., *Minds, Brains and Programs*, in *Behavioral and Brain Sciences*, 1980, pp. 417–57

unica dai processi neurobiologici che la contraddistinguono, bensì realizzare dei sistemi in grado di esibire un comportamento intelligente.[20] Pertanto al momento sembra improbabile la possibilità di realizzare una macchina dotata di coscienza proprio perché quest'ultimo è un concetto ben distinto da quello di intelligenza, anche se tendiamo a confondere i due concetti poiché nell'essere umano sono strettamente interconnessi[21] : la coscienza implica l'abilità di sentire emozioni come la gioia, il dolore, la rabbia; l'intelligenza attiene ,diversamente, alla capacità di risolvere problemi.[22]

1.2 La nozione di Algoritmo

Dopo aver delineato il concetto di Intelligenza Artificiale, si ritiene opportuno approfondire la nozione di 'algoritmo'.

Sebbene il concetto di algoritmo abbia oltre mille anni[23], ha assunto particolare rilievo all'inizio di questo secolo grazie alla rivoluzione culturale e tecnologica che ha promosso in ogni campo. Si pensi all'utilizzo dei GPS per la geolocalizzazione [24], al riconoscimento di persone, animali e cose presenti all'interno di un'immagine, alla lettura dei dati biometrici (es. volto, iride) e ai veicoli a guida autonoma.

[20] A. C. Varzi, *L'intelligenza e l'artificiale*, in *KOS. Rivista di Scienza e Etica*, 1991, pp. 12–19.
[21] Y.N. Harari, *21 Lessons for the 21th Century*, Jonathan Cape, Londra, 2018. L'autore fa l'esempio di un aeroplano che, pur non essendo dotato di penne come gli uccelli, riesce comunque a volare più veloce. Allo stesso modo i computer, pur non essendo dotati di emozioni, riescono a risolvere problemi estremamente complessi in maniera efficiente.
[22] E. Sadin, *ult. op. cit.*
[23] Il termine è stato coniato nel IX secolo dal matematico Muhammad Ibn Musa.
[24] I sistemi di calcolo del percorso ottimale sfruttano gli algoritmi per efficientare il percorso in termini di tempo, kilometri percorsi e consumo di carburante, ricalcolando il tragitto anche in relazione ad elementi contingenti come traffico e/o lavori su strada.

Recentemente gli algoritmi sono stati applicati anche in campi quali la medicina [25], l'agricoltura di precisione[26] e l'intrattenimento.[27]

In termini generali, con algoritmo si esprime il concetto di procedura generale, di metodo sistematico valido per la risoluzione di una certa classe di problemi[28]. Volendo utilizzare una nozione più tecnica, è definibile come una sequenza finita e ordinata di regole (istruzioni e operazioni) applicate al fine di ottenere un risultato in un tempo finito. Tale sequenza di regole può essere parte di un processo automatizzato di esecuzione o può avvalersi di modelli messi a punto grazie all'apprendimento automatico[29].

Gli algoritmi possono eseguire operazioni di calcolo, di elaborazione dati e ragionamento automatizzato[30] che sarebbero estremamente complesse per un essere umano sulla base di un insieme di dati in ingresso (*input*) e producendo dei dati in uscita (*output*), operando con il potere della *riduzione computazionale* di alcuni elementi; lo scopo perseguito è rendere tali elementi calcolabili, quantificabili e misurabili tramite un linguaggio matematico[31].

Il procedimento deve rispettare una serie di proprietà definitorie: deve essere *generale* (cioè deve essere valido non per un unico problema, ma per tutti i problemi appartenenti alla specifica classe considerata), *effettivo,* deve cioè essere effettivamente eseguibile da un esecutore umano o meccanico (le frasi di questo linguaggio si dicono "istruzioni"), *non ambiguo* (le istruzioni devono essere fornite

[25] Nel campo della medicina gli algoritmi predittivi sono utilizzati per effettuare prognosi e diagnosi precoci, ad esempio per il riconoscimento di patologie dalle radiografie (c.d *image recognition*) o per prevedere mutazioni genetiche.
[26] L'agricoltura di precisione sfrutta i dati delle previsioni metereologiche per individuare l'esatta quantità di acqua di cui necessitano le coltivazioni.
[27] Si pensi al meccanismo di *Recommandation System* sfruttato dai sitit di e- commerce come Amazon o da piattaforme come Spotify, Netflix e Prime Video che, sulla base degli interessi dell'utente, suggerisce il contenuto che più si adatta alle preferenze di quest'utlimo.
[28] G.Lazzari, *L'enciclopedia Treccani,* Napoli, 1977 per la definizione di algoritmo.
[29] CEPEJ*, Carta etica sull'uso dell'intelligenza artificiale nei sistemi giudiziari e nel loro ambiente*, 2018
[30] Vedere sul punto anche A. Vespignani*, L'algoritmo e l'oracolo: come la scienza predice il futuro e ci aiuta a cambiarlo,* Il Saggiatore, Milano, 2019
[31] B. Romano, *Algoritmi al potere: calcolo, giudizio, pensiero*, Giappichelli, Torino, 2018, p. 8

in un linguaggio comprensibile all'esecutore), *dettagliato* (le condizioni iniziali e finali devono essere precise e puntuali, inoltre deve essere specificato l'ordine d'esecuzione) e *finito* (le istruzioni devono essere limitate e in una successione finita; l'esecuzione deve avvenire in un tempo contenuto).[32]

Pur non essendovi ancora accordo sulla classificazione degli algoritmi, si possono individuare quattro macroaree di applicazione [33]:

1. algoritmi di *ordinamento* per la creazione di elenchi ordinando un numero smisurato di scelte[34];
2. algoritmi di *classificazione* in grado di raggruppare i dati per categorie[35];
3. algoritmi di *associazione* per ricercare le relazioni esistenti tra gli oggetti[36];
4. algoritmi *filtro* per isolare le informazioni più rilevanti[37].

La maggior parte degli algoritmi è in grado di combinare queste quattro funzioni riuscendo a sfruttare diversi meccanismi.

Quelli maggiormente utilizzati sono due: la programmazione diretta da parte dell'essere umano e il *machine learning* (si fornisce alla macchina un obiettivo, dei dati e un *feedback* che ci si aspetta e sarà lei a individuare il modo più efficiente di raggiungere l'obiettivo). Entrambi i metodi comportano vantaggi e svantaggi: il primo implica la possibilità di risolvere solo un numero limitato di problemi, quelli cioè che potrebbe risolvere anche l'uomo inserendo una serie di istruzioni, con tutti i rischi che l'intervento umano comporta (ad esempio il rischio che l'algoritmo sia programmato con dei pregiudizi culturali), mentre il secondo metodo risolverà il problema scegliendo una strada che potrebbe risultare indecifrabile anche per i programmatori più esperti e imprevedibile nella sua evoluzione (con evidenti

[32] E. Peres, *Che cosa sono gli algoritmi*, Salani Editore, Milano, 2020, pp. 3 et seq.
[33] Berryhill, Jamie, Kévin Kok Heang, Rob Clogher and K. McBride. "*Hello, World: Artificial intelligence and its use in the public sector.*" (2019), p. 17
[34] Si pensi al suggerimento del percorso più breve da parte del navigatore
[35] A titolo di esempio, sono in grado di analizzare il contenuto di video e rimuovere i contenuti inappropriati.
[36] Ad esempio, gli algoritmi delle app di incontri che cercano di individuare punti di contatto tra gli iscritti.
[37] Si pensi a Siri, che deve filtrare la nostra voce distinguendola dai restanti rumori ambientali.

criticità in termini di trasparenza e accessibilità al meccanismo decisionale alla base del risultato ottenuto)[38].

1.2.1 Il Machine learning

Come disciplina scientifica, l'Intelligenza Artificiale (da qui AI) comprende diversi approcci e diverse tecniche, come l'apprendimento automatico (di cui l'apprendimento profondo e l'apprendimento per rinforzo sono esempi specifici), il ragionamento meccanico (che include la pianificazione, la programmazione, la rappresentazione delle conoscenze e il ragionamento, la ricerca e l'ottimizzazione) e la robotica (che comprende il controllo, la percezione, i sensori e gli attuatori e l'integrazione di tutte le altre tecniche nei sistemi cyberfisici) [39].

Nello specifico, il *machine learning* [40] (o "apprendimento automatico") è una particolare tecnica di apprendimento ispirata a quella propria degli esseri umani, tramite cioè induzione di principi generali a partire dall'osservazione di dati con la possibilità di imparare nuove funzioni senza essere esplicitamente programmati per farlo[41], proprio perchè dotati di *autonomia funzionale.*

L'obiettivo è consentire all'algoritmo di apprendere da solo, identificando le relazioni nei dati osservati senza servirsi di regole o modelli forniti di volta in volta dal programmatore [42]. In questo modo, tramite l'esperienza, il sistema evolve e migliora le proprie prestazioni senza che si renda necessario il sistematico intervento dell'uomo, riuscendo a elaborare nuovi algoritmi in autonomia.

---

[38] *Supra* nota 33

[39] High-Level Expert Group on AI (2019). *Ethics guidelines for trustworthy AI* (Report). European Commission.

[40] Il termine è stato utilizzato per la prima volta da Arthur Samuel nel 1959 nell'articolo "*Alcuni studi sul machine learning usando il gioco della dama".* Il concetto deve essere distinto da quello di intelligenza artificiale in quanto il *machine learning* costituisce una modalità di apprendimento specifico utilizzato nell'ambito della prima, di cui rappresenta un sottoinsieme.

[41] S. Quintarelli, *ult. op. cit.,* p. 30

[42] A. Vespignani*,ult. op. cit.*

Il vantaggio è che in questo modo il sistema riesce ad adattarsi a nuove circostanze non previste dallo sviluppatore al momento della programmazione, rilevando *pattern* in qualsivoglia tipo di dati, creando nuovi comportamenti per adattarsi alle richieste ricevute e prendendo decisioni in base al successo o fallimento delle scelte precedenti[43].

La macchina può apprendere sulla base di tre modelli principali:

1. l'apprendimento supervisionato da un essere umano (nel quale l'algoritmo genera classi secondo modelli forniti dal programmatore e dotati dell'etichetta corretta[44]);
2. l'apprendimento non supervisionato (dove la classificazione è creata autonomamente dall'algoritmo [45] , viene impiegato nelle operazioni di raggruppamento dei dati);
3. l'apprendimento per rinforzo (in cui l'algoritmo impara lavorando per tentativi e cercando lo schema che minimizza i suoi errori tramite un meccanismo di punizione-ricompensa[46]).

Generalmente alla base dell'architettura di questi algoritmi vi sono il *deep learning*[47] e le reti neurali[48]. Un esempio del funzionamento di tale meccanismo è fornito dalla cosiddetta "pubblicità tracciante", tramite la quale l'algoritmo presenta proposte pubblicitarie strettamente connesse agli interessi dell'individuo

[43]Il termine "apprendere" non va inteso nel senso umano del termine: la macchina non è dotata di funzioni intellettive, riesce tuttavia a collezione un numero crescente di informazioni e rielaborarle per aumentare le proprie prestaizoni.

[44] R. Marmo, *Algoritmi per l'intelligenza artificiale. Progettazione dell'algoritmo - Dati e Machine Learning - Neural Network - Deep Learning,* Hoepli, Milano, 2020, Capitolo 8

[45] *Ibidem*

[46] A. Vespignani*, ult. op. cit.*

[47]Il *deep learning* costituisce un sottoinsieme del machine learning: è un metodo di apprendimento automatico che utilizza dei modelli di reti neurali estremamente complessi, composti da numerosi livelli che combinano differenti algoritmi.

[48] Le reti neurali sono sistemi informatici ispirati alle reti neurali biologiche che costituiscono il cervello dei mammiferi; tali sistemi apprendono lo svolgimento di determinati compiti prendendo in considerazione degli esempi. Possono riuscire a imparare a identificare immagini in cui figurano dei gatti senza avere una preliminare conoscenza di tali animali semplicemente analizzando esempi di immagini etichettate manualmente come "gatto". Da CEPEJ*, Carta etica sull'uso dell'intelligenza artificiale nei sistemi giudiziari e nel loro ambiente*, 2018

analizzando le ricerche effettuate in rete dallo stesso[49] tramite una stringa definita *cookie*: la funzione "prodotti consigliati" sui siti di *e-commerce* come Amazon sfruttano proprio questo approccio.

Al netto di tali considerazioni, le applicazioni sono ben più numerose: gli algoritmi sono alla base della diagnostica per immagini in medicina, dei suggerimenti musicali di piattaforme come Spotify e di film e serie tv come Netflix e svolgono un ruolo centrale anche per l'uso degli assistenti vocali come Siri e Alexa, che sfruttano il *machine learning* per comprendere il linguaggio umano e migliorare le proprie prestazioni[50]. Un altro sistema che si basa su algoritmi di *machine learning* è il navigatore GPS: l'algoritmo rappresenta il problema del percorso più breve in una struttura composta da nodi (città, incroci, luoghi) uniti da archi (i collegamenti tra queste località): l'algoritmo sarà in grado di proporci la soluzione che massimizza il concetto di ottimalità[51].

Il ruolo centrale nel funzionamento dell'apprendimento automatico lo svolgono i *Big Data*, grandi insieme di dati raccolti su larga scala provenienti da fonte eterologhe e organizzati in banche dati[52]. I dati svolgono un ruolo chiave nella prima fase di funzionamento dell'algoritmo, detta "fase di allenamento", in cui l'algoritmo legge un insieme di dati esistenti: maggiore è la quantità di dati a disposizione, maggiore è la capacità dell'algoritmo di derivare una funzione complessa per descriverli in modo da saper riconoscere scenari nuovi, generalizzando a situazioni sconosciute[53].

Nonostante i numerosi vantaggi apportati, l'impiego degli algoritmi di *machine learning* ha sollevato diverse criticità: la preoccupazione primaria sta nel fatto che in molti casi non saranno comprensibili e interpretabili per l'essere umano le ragioni che hanno portato l'algoritmo a quel risultato e il peso attribuito ai singoli fattori da valutare (si parla di *machine learning explicability*), aspetto fondamentale per

[49] D. Talia, *ult. op. cit.* vedere sul punto il capitolo "Big Data e società calcolabile".
[50] S. Quintarelli. *ult. op. cit,* p. 47
[51] S. Quintarelli, *ult. op. cit.* p. 40
[52] CEPEJ*, Carta etica sull'uso dell'intelligenza artificiale nei sistemi giudiziari e nel loro ambiente*, 2018
[53] *Supra* nota 51 (Quintarelli)

individuare e correggere eventuali errori che potrebbero condurre a prendere decisioni inique. Il problema è anche nella qualità dei dati immessi su cui lavora la macchina, i quali, come si esporrà nel corso della trattazione, potrebbero nascondere dei pregiudizi culturali[54].

1.2.2 Decisione automatizzata e crescente impiego nei processi decisionali quotidiani

L'impatto degli algoritmi sulla vita quotidiana è talmente complesso e profondo da poter configurare una possibile trasformazione antropologica dell'*humanitas*[55].

L'utilità è innegabile, tuttavia non bisogna attribuire loro un potere dogmatico: pur processando una quantità impensabile di dati per le capacità umane, gli algoritmi non costituiscono e non esauriscono le peculiarità della condizione umana, che risiede nell' abilità di eccedere i dati per avere una visione più ampia e completa del fenomeno considerato.[56] Il rischio è che da semplici strumenti deputati ad assisterci nelle nostre attività, gli algoritmi diventino entità che plasmano le nostre vite e a cui deleghiamo le decisioni nella convinzione che le scelte operate siano oggettivamente migliori delle nostre[57].

Come evidenziato da Eric Sadin nel libro *Critica alla ragione artificiale,* oggi più una decisione si rivela complicata, più sorge la necessità di rivolgersi a sistemi automatici capaci di istruire l'azione umana. Un caso eclatante e particolarmente critico è rappresentato dall'impiego di algoritmi dai giudici nei tribunali statunitensi per valutare il rischio di recidiva di un accusato nella convinzione che tali programmi siano in grado di enunciare una verità oggettiva e indubitabile (a cui spesso si perviene tramite processi che spesso neanche i più esperti programmatori sono in grado di comprendere).

---

[54] R. Marmo, *ult. op. cit.*
[55] Come ha affermato Bruno Romano, Professore di Filosofia del diritto presso l'Università La Sapienza di Roma
[56] B. Romano, *ult. op. cit.* p. 62
[57] E. Sadin, *ult. op. cit.,* p. 73

Secondo Sadin, si può parlare di un rifiuto della vulnerabilità intrinseca dell'uomo che sta producendo una rivoluzione: gli algoritmi, da mero strumento al servizio delle attività umane, arrivano così a plasmare la nostra realtà erodendo le nostre capacità di giudizio e azione. Essi si fanno carico dei nostri bisogni, arrivando a conoscere le nostre emozioni e i nostri desideri[58].

In questo modo le analisi dei dati diverranno il nuovo metro dell'obiettività, come afferma Harari: "è stata creata una nuova grande narrazione universale che legittima l'autorità di algoritmi e *big data*" dando vita a una nuova convinzione chiamata *dataismo*[59]. L'alleanza tra la rivoluzione biotecnologica e quella della informazione tecnologica produrranno algoritmi che saranno in grado di capire le nostre emozioni meglio di noi stessi e ciò che prima era inaccessibile a terzi diverrà comprensibile e appannaggio di modelli matematici[60], comportando numerosi vantaggi[61] ma anche notevoli criticità[62].

Come rilevato nel paragrafo precedente, gli algoritmi sono impiegati ormai in modo più o meno consapevole nei nostri processi decisionali quotidiani: dall'ammontare del finanziamento che può esserci concesso al percorso più breve per arrivare a destinazione, sino alle applicazioni più controverse nell'ambito della giustizia[63]. Ad esempio, lo stesso suggerimento di un film da vedere è regolato dall'azione di algoritmi che, sulla base dei film precedentemente visti, attinge ad un database e calcola statisticamente il prodotto che può soddisfare maggiormente i nostri gusti[64]. Altre applicazioni possono essere rinvenute nel campo della medicina

---

[58] *Supra* nota 57, pp. 156-157 (Sadin)
[59] A. Carleo (a cura di), *Decisione robotica*, Il Mulino, Bologna, 2019
[60] Y. N. Harari*, ult. op. cit.*
[61] Harari immagina un futuro prossimo in cui gli algoritmi riusciranno a captare l'insorgere di malattie come l'influenza o l'Alzhaimer ancor prima che l'uomo possa percepire i primi sintomi.
[62] Prima fra tutte il fatto che gli algoritmi non sono perfetti e necessariamente infallibili: sono pur sempre programmati dall'essere umano che può anche inconsapevolmente progettarli con un pregiudizio culturale. Inoltre, possono non tener conto di variabili fondamentali e i meccanismi di funzionamento e il peso dato ai singoli elementi valutati non sempre è comprensibile.
[63] Negli USA vengono utilizzati algoritmi come COMPASS per calcolare il rischio di recidiva dell'accusato; l'algoritmo si basa sulla compilazione di un questionario e poi fornisce una risposta sulla base di calcoli statistici.
[64] *Supra* nota 60 (Harari)

diagnostica, dell'istruzione (per la valutazione, prognosi e orientamento professionale), dei trasporti (basti pensare ai veicoli a guida autonoma), delle assicurazioni (per stimare e analizzare il profilo di rischio), del marketing[65] e dell'agricoltura (per predire il momento ideale della semina e calcolare l'indice della potenziale acqua richiesta dal raccolto[66]). Persino nel mondo dell'arte, fino a poco tempo fa considerato ad esclusivo appannaggio dell'essere umano e della sua sensibilità, si sono sviluppati algoritmi in grado di riprodurre fedelmente stili di pittori e musicisti.

Numerose perplessità solleva l'applicazione dei suddetti algoritmi in campi più delicati come la giustizia e la prevenzione dei crimini. In particolare, recentemente in Inghilterra è stato introdotto l'algoritmo HART (e negli Stati Uniti PredPol) per individuare dove e quando saranno compiuti determinati tipi di crimini basandosi sui rapporti di polizia e dati statistici[67] (con il rischio di un'eccessiva sorveglianza da parte della polizia dei quartieri considerati più a rischio). Anche i sistemi di video-sorveglianza sono dotati di software in grado di analizzare anomalie comportamentali nei luoghi pubblici e segnalare la commissione di crimini[68].

Anche in Italia si utilizzano degli algoritmi come *Key Crime* e *XLaw* che coadiuvano le forze di polizia nella prevenzione e controllo del territorio. In USA si utilizza l'algoritmo di valutazione del rischio[69] per calcolare la possibilità di recidiva dell'accusato e fornire informazioni utili per il giudice alla definizione della sentenza o della misura cautelare da applicare.

Gli algoritmi che utilizzano i big data nei diversi settori esaminati come la pubblicità online o la condanna penale possono rendere le decisioni più obiettive e

[65] D. De Kerckhove, *La decisione datacratica,* in *Decisione Robotica ult. op. cit.,* pp. 102-103
[66] S. Quintarelli, *ult. op. cit.,* pp. 54-56
[67] *Infra* capitolo III, par. 4
[68] *Ibidem*
[69] *Infra* capitolo III, par. 5

guidate dai dati ma non evitano il perpetrarsi delle disuguaglianze sociali e della discriminazione[70].

Oltre a problemi di iniquità, sarà penalizzata anche la nostra autonomia (minore libertà di decisione e maggiore dipendenza dalle macchine), la proprietà individuale di pensiero (l'intelligenza artificiale è in grado di rintracciare i nostri pensieri, desideri e emozioni) e comporterà una diminuzione della nostra interiorità psicologica: la macchina acquisisce le funzioni cognitive peculiari (memoria, intelligenza, razionalità)[71].

[70] Un recente studio ha per altro dimostrato come la funzione di Amazon "consegna in giornata" non fosse disponibile nei quartieri neri per ragioni che le aziende non erano in grado di spiegare. Un altro studio ha evidenziato come gli annunci per posti di lavoro ad alto reddito pubblicizzati da Google sarebbero stati indirizzati in misura minore alle donne.
[71] *Supra* nota 70, pp. 77 et ss.

CAPITOLO II

IL QUADRO GIURIDICO DI RIFERIMENTO IN MATERIA DI INTELLIGENZA ARTIFICIALE

2.1 Strumenti giuridici internazionali in materia di intelligenza artificiale

Negli ultimi anni è sorto un acceso dibattito in ambito internazionale in materia di intelligenza artificiale, proprio in virtù delle straordinarie opportunità ma anche dei numerosi rischi (quali meccanismi decisionali opachi, discriminazioni basate sul genere o di altro tipo e pregiudizi culturali[72]) connessi all'impiego diffuso in tutti i settori sociali ed economici di tale tecnologia. È dunque sorta la conseguente necessità di regolare il fenomeno dettando una serie di principi e linee guida per garantire un impiego di tale tecnologia nel pieno rispetto dei diritti umani sanciti universalmente[73].

Lo strumento che meglio è riuscito ad adattarsi alle necessità di regolazione del fenomeno è stato quello del s*oft law*[74] che, oltre a fornire codici di condotta e linee guida da seguire nell'impiego responsabile di queste tecnologie, ha anche il vantaggio di garantire flessibilità per adattarsi al meglio ai rapidi sviluppi nel campo[75]. Già nel 2015, conscio dei rapidi sviluppi che a breve si sarebbero verificati in materia, l'*United Nations Interregional Crime and Justice Research Institute* (UNICRI) ha istituito il Centro sull'Intelligenza artificiale e la robotica "*to help*

[72] È cruciale ricordare che gli algoritmi alla base dell'IA sono programmati da esseri umani che potrebbero volontariamente o involontariamente programmarli con tali pregiudizi che contribuirebbero a cristallizzare situazioni di iniquità sociale.
[73] Come ha affermato il segretario generale di Amnesty International Salil Shetty «AI is built by humans and it will be shaped by human values. If we build AI systems that are a mirror to our current societies, they will be riddled with the historical biases and inequalities of our societies.»
[74] Per strumenti di *soft law* si intendono atti privi di efficacia vincolante.
, p. 5

focus expertise on Artificial Intelligence (AI) throughout the UN in a single agency.".

L'High Commissioner for Human Rights (UNCHR) dell'ONU ha emanato numerosi rapporti aventi ad oggetto le implicazioni dell'intelligenza artificiale su alcuni diritti umani, con particolare riguardo ai problemi di discriminazione che possono derivare dall'impiego di tali tecnologie[76]. Altri atti emanati in materia, particolarmente incisivi sono "*Le linee guida intergovernative sulle politiche relative all'intelligenza artificiale"* [77] da parte dell'OCSE e le "*Guidance for Regulation of Artificial Intelligence Application"*[78]*,* emanate nel 2020 da parte del governo americano.

Ai fini della presente analisi si analizzerà un atto particolarmente significativo: la dichiarazione di Toronto per la protezione del diritto all'uguaglianza e alla non discriminazione nei sistemi di *machine learning.*

2.1.1 La dichiarazione di Toronto: la protezione del diritto all'uguaglianza e alla non discriminazione nei sistemi di machine learning

La Dichiarazione di Toronto[79], adottata nel Maggio 2018, è un documento che esamina l'impatto potenzialmente lesivo dei sistemi di *machine learning* sui diritti umani, con particolare riferimento al diritto all'uguaglianza e alla non discriminazione. Assieme alla Dichiarazione di Montreal per uno sviluppo

---

[76] *Promotion, protection and enjoyment of human rights on the Internet: ways to bridge the gender digital divide from a human rights perspective* - Report of the United Nations High Commissioner for Human Rights- A/HRC/35/9

[77] Nel documento in cui si affermano cinque principi: l'IA dovrebbe essere utilizzata per sostenere una crescita inclusiva, dovrebbe rispettare lo stato di diritto e i diritti umani, dovrebbe essere trasparente e accessibile, sicura per tutta la durata del suo utilizzo e i soggetti che sviluppano i sistemi di IA dovrebbero essere ritenuti responsabili del loro corretto funzionamento.

[78] Viene incoraggiato l'impiego dei sistemi di IA ma allo stesso tempo si richiede di osservare i principi di equità e non discriminazione poiché vi è il rischio che l'applicazione "introduca pregiudizi culturali che producono risultati discriminatori o decisioni che minano la fiducia verso l'IA". Si richiedono garanzie in termini di accessibilità, trasparenza (per accedere ai meccanismi di funzionamento tramite i quali l'IA raggiunge un determinato risultato) e sicurezza nell'uso e sviluppo dell'IA.

[79] Su proposta, in particolare, di Amnesty International e Access Now.

responsabile dell'Intelligenza artificiale[80] contribuisce a definire un quadro etico per guidare lo sviluppo e l'utilizzo dell'AI.

Nel preambolo della dichiarazione si riconosce il potenziale dei sistemi di *machine learning* per la promozione dello sviluppo economico e sociale ma si pone l'accento anche sui rischi implicati nell'uso di essi; si legge infatti: "*Machine learning systems, which can be opaque and include unexplainable processes, can contribute to discriminatory or otherwise repressive practices if adopted and implemented without necessary safeguards*"[81].

Dopo aver richiamato la responsabilità degli Stati a promuovere e proteggere i diritti umani, la dichiarazione prende in esame la tutela del diritto all'uguaglianza e alla non discriminazione: vi è infatti il rischio che "*Existing patterns of structural discrimination may be reproduced and aggravated in situations that are particular to these technologies system goals that create self-fulfilling markers of success and reinforce patterns of inequality, or issues arising from using non-representative or biased datasets"*[82].

Particolare attenzione è rivolta al rischio dei pregiudizi che possono nascondere i meccanismi di *machine learning*: i programmatori degli algoritmi potrebbero proiettare i loro *bias* culturali (anche inconsapevolmente) contribuendo a perpetrare e cristallizzare discriminazioni verso alcuni gruppi in termini di razza, cultura, sesso, età e condizioni socio-economiche.[83] Certe proiezioni di valori potrebbero nascondersi sia all'interno del codice del programma, sia nei dati che vengono forniti al programma stesso (che potrebbero rispecchiare le convinzioni di chi li ha forniti), rendendo difficile individuare e neutralizzare rischi di discriminazione[84].

[80] Elaborata dall' Università di Montréal, fissa dieci principi etici da osservare per promuovere gli interessi fondamentali delle persone e dei gruppi sociali.
[81] Vedere *The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems*
[82] *Supra* nota 81, par. 16
[83] *Supra* nota 81
[84] *Infra* Capitolo IV, par. 4.1.

Si è portati a ritenere che il risultato dato dall'algoritmo sia necessariamente neutrale e imparziale proprio perchè proveniente da una macchina, dimenticando che quella stessa macchina viene programmata da un essere umano.

Al fine di mitigare e ridurre l'impatto lesivo di tali tecnologie sui diritti umani, secondo la Dichiarazione gli Stati devono:

I. *Identificare i rischi connessi all'impiego dei sistemi di machine learning*, conducendo valutazioni d'impatto periodiche per identificare fonti di rischio di risultati discriminatori o lesivi (ad esempio nella fase della progettazione e supervisione degli algoritmi e nell'elaborazione dei dati), sottoponendo i sistemi a test periodici per rilevare eventuali *bias* impliciti o "*feedback loops*"[85].

II. *Assicurare trasparenza e responsabilità* in caso di danno: ciò include diritto di accesso all'algoritmo, esplicabilità dei processi alla base del loro funzionamento tramite una politica di *public disclosure,* permettendo a esperti indipendenti di effettuare controlli e valutazioni.

III. Rinforzare i meccanismi di supervisione sensibilizzando circa i rischi di discriminazione e gli altri effetti lesivi connessi, garantendo un'adeguata formazione anche a coloro che utilizzano tali sistemi nei vari settori e creando meccanismi di controllo indipendente, anche da parte di autorità giudiziarie. È fondamentale, inoltre, garantire che le decisioni supportate dal *machine learning* rispettino gli standard universali per un processo equo, visto l'impiego che se ne fa anche in ambito giudiziario.

IV. Promuovere l'eguaglianza tramite programmi inclusivi che valorizzino la diversità e l'equità in diversi settori.

[85] *Infra* capitolo IV, par. 4.3.

Tali obblighi devono essere osservati anche quando gli Stati si affidano a contraenti privati per progettare queste tecnologie e utilizzarle in un contesto pubblico, non potendo rinunciare ai loro obblighi in materia di prevenzione alla discriminazione e responsabilità di riparazione per i danni ai diritti umani nel fornire suddetti servizi.[86] Inoltre, gli Stati hanno l'obbligo di garantire il rispetto dei diritti umani anche quando gli attori in gioco sono privati[87] nel rispetto del principio di *due diligence*[88], identificando potenziali risultati discriminatori in anticipo, prendendo azioni effettive per prevenire e rimuovere gli effetti discriminatori e assicurando la trasparenza di tali sistemi. Infine, la Dichiarazione richiede che siano messi appunto *rimedi effettivi*[89] cui fare ricorso in caso di lesione dei propri diritti da parte dei sistemi decisionali automatizzati ("*victims of human rights violations or abuses must have access to prompt and effective remedies, and those responsible for the violations must be held to accoun*t"[90]).

[86] *Supra* nota 81, par. 84

[87] «*States parties must therefore adopt measures, which should include legislation, to ensure that individuals and entities in the private sphere do not discriminate on prohibited grounds*»

[88] In base al quale lo Stato deve prevenire e reprimere violazioni dei diritti umani laddove ne abbia conoscenza e possibilità in relazione a condotte assunte dai privati. Vedere sul punto Council of Europe's Recommendation CM/Rec (2018) 2 of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries.

[89] L'accesso a rimedi effettivi potrebbe essere messo a rischio dall'*opacità* degli algoritmi, che ostacolerebbero così il diritto di difesa (non potendo conoscere il meccanismo di funzionamento dell'algoritmo né il peso dato ai singoli elementi valutati dallo stesso il ricorrente non ha a disposizione elementi essenziali), contribuendo a determinare una asimmetria informativa tra le parti. La criticità è particolarmente acuta quando tali sistemi sono impiegati per prendere decisioni all'interno dei sistemi giudiziari: è necessario garantire il rispetto dei principi di equo processo e chiarire chi è legalmente responsabile per le decisioni prese tramite tali sistemi, garantendo una riparazione adeguata a chi ha subito una violazione e adeguate garanzie di non ripetizione.

[90] *Supra* nota 81, par. 52

2.2 Strumenti giuridici a livello europeo

La "quarta rivoluzione" ha ormai avviato un processo di trasformazione non solo a livello antropologico, ma anche giuridico: il diritto risente dei profondi cambiamenti della società e la necessità di regolazione del fenomeno inizia a essere avvertita in tutto il mondo.

L'Europa si è per prima occupata del problema della regolamentazione dell'AI, intuendo le potenziali minacce che avrebbe potuto comportare per i diritti umani e contribuendo a influenzare significativamente il successivo dibattito internazionale grazie ai numerosi atti emanati in materia.

Già nel 2016 con il GPDR in materia di protezione dei dati personali erano state affrontate alcune criticità legate all'impiego di meccanismi decisionali automatizzati di *machine learning*.

Successivamente la strategia europea per la regolazione dell'AI è stata implementata con la presentazione da parte della Commissione europea del documento *"L'intelligenza artificiale per l'Europa"*[91] nell'Aprile del 2018. È stato poi pubblicato il "*Piano coordinato sull'intelligenza artificiale*[92]*"* ed istituito un Gruppo di esperti al fine di promuovere un AI trasparente e affidabile, culminato con l'adozione da parte del Consiglio d'Europa della "*Carta etica europea sull'uso dell'Intelligenza artificiale*" e delle *"Ethics Guidelines for trustworthy AI*[93]".

Sull'onda di tali interventi, a titolo di esempio, si riporta l'esperienza dell'Inghilterra, che ha emanato un atto significativo in materia, noto come "*Data*

[91] Nel documento si fornisce una definizione di intelligenza artificiale e si espone l'iniziativa europea a dare impulso all'adozione dell'IA in numerosi settori per implementare lo sviluppo tecnologico, assicurando al tempo stesso un quadro etico e giuridico adeguato coerente con la carta EDU. Sono inoltre individuati sette requisiti fondamentali di cui deve essere dotata: sette requisiti fondamentali individuati negli orientamenti del gruppo di esperti ad alto livello: intervento e sorveglianza umani, robustezza tecnica e sicurezza, riservatezza e governance dei dati, trasparenza, diversità, non discriminazione ed equità, benessere sociale e ambientale, e accountability.

[92] Predisposto insieme agli Stati membri per promuovere lo sviluppo e l'utilizzo dell'intelligenza artificiale in Europa.

[93] In cui si elencano sette requisiti chiave per una intelligenza artificiale affidabile: sorveglianza umana, robustezza e sicurezza, privacy e gestione dei dati, trasparenza, diversità e non discriminazione, benessere sociale e ambientale e responsabilità.

Protection Act". La Sezione 14 è dedicata alle decisioni automatizzate autorizzate dalla legge, in correlazione con l'art. 22 del GPDR: nel caso in cui la decisione sia stata presa basandosi esclusivamente su un processo decisionale automatizzato, è riconosciuto al soggetto una tutela in termini di notificazione all'interessato e il diritto di richiedere che la decisione venga riconsiderata o che ne sia presa una nuova che non sia basata unicamente su un processo automatizzato[94]. L'Italia ha recentemente adottato una Strategia nazionale per l'Intelligenza Artificiale[95], elaborata dal Ministero dello Sviluppo Economico sulla base delle proposte definite dal gruppo di esperti selezionati dal MISE, per massimizzare i benefici derivanti dalla trasformazione digitale e tecnologica.

I Paesi Bassi hanno invece adottato nel 2019 lo *"Strategic Action Plan for Artificial Intelligence"* contenente numerose iniziative per assicurare uno sviluppo adeguato dell'AI e allo stesso tempo garantire la protezione dei valori e diritti umani fondamentali[96].

Nel Febbraio 2020 è stato infine adottato da parte della Commissione europea il *"Libro bianco sull'intelligenza artificiale",* in cui sono definiti approcci normativi e orientamenti che hanno come obiettivo la predisposizione di una strategia per allineare gli sforzi a livello europeo, nazionale e regionale nello sviluppo sicuro e affidabile dell'AI in Europa, nel rispetto dei valori e dei diritti dei cittadini UE per creare un "ecosistema di eccellenza e fiducia" dove i diritti fondamentali dei cittadini siano garantiti anche limitando le applicazioni considerate ad alto rischio[97].

---

[94] «*a) the controller must, as soon as reasonably practicable, notify the data subject in writing that a decision has been taken based solely on automated processing, and b) the data subject may, before the end of the period of 1 month beginning with receipt of the notification, request the controller to i) reconsider the decision, or ii) take a new decision that is not based solely on automated processing*.»

[95] Consultabile sul sito del Ministero all'indirizzo https://www.mise.gov.it/index.php/it/per-i-media/notizie/2041503-intelligenza-artificiale-al-via-la-consultazione-pubblica-sulla-strategia-nazionale

[96] M. Spielkamp, *Automating Society: Taking Stock of Automated Decision-Making in the EU,* BertelsmannStiftung Studies 2019, p.56 in www.algorithmwatch.org/automating-society

[97] E' considerata ad alto rischio se applicata a settori "sensibili" (come l'assistenza sanitaria, il settore giudiziario, le pratiche di assunzione) e se l'uso nel settore in questione può generare rischi significativi). Sul punto vedere il documento adottato dalla

Con la legge 22 aprile 2021 n. 6 l'Italia ha ratificato il Protocollo alla Convenzione n. 108 [98] sulla protezione delle persone rispetto al trattamento automatizzato di dati a carattere personale, che sancisce ulteriori tutele nei confronti dei dati utilizzati nelle procedure algoritmiche.

2.2.1 Il Regolamento generale della protezione dei dati (GDPR n. 2016/679)

L'utilizzo di algoritmi solleva la questione della protezione dei dati personali trattati: a livello europeo si è avvertita l'esigenza di mettere in atto politiche di prevenzione finalizzate a neutralizzare i potenziali rischi associati all'utilizzo dei dati trattati da tali algoritmi e alle conseguenze del loro utilizzo per le persone e la società in generale [99]. Le norme più rilevanti ai fini di tutela dell'individuo nell'ambito di decisioni automatizzate sono ricavabili principalmente dalla disciplina europea in materia di trattamento dei dati personali, che vuole creare così una barriera simbolica alle decisioni prese senza intervento umano. Le principali norme di riferimento sono rappresentate dagli articoli 13 e 15 del GDPR in materia di trasparenza algoritmica, che stabiliscono il diritto dell'interessato a conoscere l'esistenza di processi decisionali automatizzati che lo riguardino ed a ricevere informazioni sulla logica utilizzata. Viene inoltre enunciato il principio di non esclusività (art. 22) e il diritto di contestazione della decisione basata unicamente su un processo decisionale automatizzato, nonché il principio di non discriminazione algoritmica fissato al considerando 71 del GDPR. L'intento perseguito dal Regolamento è proprio quello di "*di arginare il rischio di trattamenti discriminatori per l'individuo che trovino la propria origine in una cieca fiducia nell'utilizzo degli algoritmi*".

Commissione europea *Libro Bianco sull'intelligenza artificiale - Un approccio europeo all'eccellenza e alla fiducia,* pp. 19 et seq.

[98] Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data. Strasbourg, 10.X.2018 consultabile all'indirizzo https://rm.coe.int/16808ac918

[99] CEPEJ,*Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi*, par. 141

Il Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio del 27 aprile 2016 avente a oggetto la protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati è stato emanato con la finalità di fornire una protezione effettiva in materia di diritti e libertà fondamentali, con particolare riguardo al diritto alla protezione dei dati personali, rispondendo alla necessità di fronteggiare le nuove sfide proposte dalla rapida evoluzione tecnologica.

All'art. 2 comma 1 si specifica che il regolamento si applica al trattamento interamente o parzialmente automatizzato di dati personali e al trattamento non automatizzato di dati personali contenuti in un archivio o destinati a figurarvi[100]. Ha assunto un ruolo chiave nel dibattito internazionale in materia di regolazione dell'intelligenza artificiale per aver sancito principi guida a tutela dei soggetti coinvolti da una decisione automatizzata, tra cui: il diritto di contestare la decisione, il principio di non discriminazione algoritmica, il principio di non esclusività, il diritto ad accedere alla logica impiegata dall'algoritmo e secondo alcuni sarebbe ravvisabile un limitato diritto alla spiegazione[101] (*right to explanation*) del risultato dell'algoritmo.

Grazie all'intelligenza artificiale siamo in grado di delegare un gran numero di compiti e decisioni per raggiungere l'obiettivo nel modo più efficiente possibile: la volontà di delegare e automatizzare compiti è il primo motore dello sviluppo dell'Intelligenza Artificiale[102]. Il problema sorge nel momento in cui a essere delegate sono anche decisioni che richiedono valutazioni più complesse, di carattere

[100] Il comma 2 dell'art 2 lett. d) individua una serie di casi di esclusione, tra cui vale la pena sottolineare l'esclusione dei trattamenti effettuati dalle autorità competenti a fini di prevenzione, indagine, accertamento o perseguimento di reati o esecuzione di sanzioni penali, incluse la salvaguardia contro minacce alla sicurezza pubblica e la prevenzione delle stesse. La disciplina europea della decisione automatizzata sarà completata con la Direttiva 680-2016, adottata il 27 aprile 2016 in materia di relativa alla protezione delle persone fisiche con riguardo al trattamento dei dati personali da parte delle autorità competenti a fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali.

[101] L'interpretazione è piuttosto dibattuta: secondo alcuni sarebbe ravvisabile nell' art. 15, secondo altri nell'art 22 pur non essendo esplicitamente menzionato, secondo un altro orientamento non sarebbe ravvisabile un simile diritto.

[102] S. Quintarelli, *ult. op.cit*., p. 78

etico- valoriale, come nel caso delle decisioni prese in ambito giudiziario. Ci sono alcuni fattori, come ad esempio l'interesse del minore nel prendere in considerazione una misura di custodia, che devono essere necessariamente soppesati, mentre altri, come ad esempio l'etnia, che invece non dovrebbero avere alcun peso[103].

Conscio di tali criticità il legislatore europeo ha sancito l'obbligo di fornire informazioni dell'eventuale esecuzione di un processo decisionale automatizzato e assicurare l'*accessibilità alla logica utilizzata dall'algoritmo*, nonché comprendere l'importanza e le conseguenze previste di tale trattamento per l'interessato in virtù dell'art .13 comma 2 lett. *f)*, 14 comma 2lett. *g)* e 15 comma 1 lett. *h)* cercando di limitare così il fenomeno dell'opacità dell'algoritmo (cosiddette *black boxes*[104]).

Al fine di scongiurare risultati discriminatori o errori commessi dal sistema decisionale, risulta fondamentale conoscere il fattore preso in considerazione, il peso attribuito e se è stato determinante per l'*outcome* dell'algoritmo, dunque se si è fatto un uso proprio di quella informazione. L'inaccessibilità alla logica sottesa si riverbera così sul piano procedurale e sostanziale: è necessario conoscere gli autori del procedimento usato per la sua elaborazione, il meccanismo decisionale, le priorità assegnate e i dati selezionati come rilevanti. Tutto ciò, al fine di certificare che gli esiti e i presupposti della decisione siano conformi alle prescrizioni e alle finalità stabilite dalla legge e affinché siano chiare – e conseguentemente sindacabili – le modalità e le regole in base alle quali esso è stato impostato[105] ed infine per garantire il diritto di difesa.

---

[103] Hessekiel, Kira, Eliot Kim, James Tierney, Jonathan Yang, and Christopher T. Bavitz. 2018. AGTech Forum Briefing Book: State Attorneys General and Artificial Intelligence, May 8-9, 2018, Harvard Law School. Berkman Klein Center for Internet & Society.

[104] Algoritmi il cui processo di funzionamento risulta non accessibile o difficilmente comprensibile; il fenomeno si verifica principalmente perchè le richieste di accesso all'algoritmo sono respinte per ragioni di tutela della proprietà intellettuale del *software*. La criticità sta nel fatto che non è possibile conoscere il meccanismo di funzionamento interno e la base di conoscenza dell'algoritmo stesso, impedendo di capire quali fattori sono stati valutati e quanto peso è stato dato a essi.

[105] G. Pesce, *Il Consiglio di Stato ed il vizio della opacità dell'algoritmo tra diritto interno e diritto sovranazionale*, in Giustizia Amministrativa, 2020.

Un altro fondamentale principio, quello di *non discriminazione algoritmica*, è stato enunciato al Considerando 71 par. 2 del GPDR secondo cui le procedure matematiche e statistiche utilizzate devono essere appropriate e sottoposte a controllo per evitare inesattezze o errori, al fine di impedire "effetti discriminatori nei confronti di persone fisiche sulla base della razza o dell'origine etnica, delle opinioni politiche, della religione o delle convinzioni personali, dell'appartenenza sindacale, dello status genetico, dello stato di salute o dell'orientamento sessuale, ovvero un trattamento che comporti misure aventi tali effetti".

Il legislatore europeo ha inoltre enunciato il *principio di non esclusivit*à, principio restrittivo alle decisioni pubbliche prese tramite algoritmi, cercando di porre una garanzia all' articolo 22, par.1, GDPR in materia di processo decisionale automatizzato, stabilendo che "*L'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici* che lo riguardano o che incida in *modo analogo*[106] *significativamente sulla sua persona*[107]".

Per "*decisione basata unicamente sul trattamento automatizzato*" si deve intendere una decisione presa senza il coinvolgimento di un essere umano che possa influenzare ed eventualmente cambiare il risultato attraverso la sua autorità o competenza[108]. Le Linee Guida predisposte dal Comitato Europeo per la protezione dei dati personali (*Working Party* art. 29) sono intervenute sul punto fornendo una interpretazione estensiva[109] della nozione onde evitare l'aggiramento del divieto: di

[106] Le linee Guida identificano l'effetto analogo in ogni conseguenza indiretta e di riflesso rispetto all'oggetto principale della decisione.
[107] Si riporta l'esempio del rifiuto automatico di una domanda di credito *online* o pratiche di assunzione elettronica senza alcun intervento umano.
[108] GDPR. Le *Guidelines* fornite dall' Article 29 Working Party ampliano ulteriormente la portata di tale definizione suggerendo che la decisione sia da ritenersi automatizzata anche allorquando sia prevista la presenza di un essere umano ma questo non abbia potere di intervento sulla decisione operata dalla macchina, limitandosi a enunciare semplicemente il risultato ottenuto.
[109] L' Article 29 Working Party (ora sostituito dall' European Data Protection Board) ha fornito una interpretazione *estensiva* della nozione stabilendo che si debba considerare decisione basata unicamente sul trattamento automatizzato una decisione in cui l'utilizzatore i) non abbia il potere di contestare o modificare la decisione ii) non possa esprimere frequentemente il proprio dissenso rispetto alle decisioni suggerite dagli strumenti di supporto (dal pdf recenti arresti della giustizia amministrativa)

conseguenza se l'intervento umano fosse solo apparente, non sfuggirebbe al divieto enunciato [110]. Pertanto, non dovrebbe comportare l'esclusione dell'intervento umano nel processo decisionale, il quale deve potersi dissociare dalla soluzione proposta all'esito del processo automatizzato[111].

Per il legislatore Europeo, l'intervento umano è visto come un "controllo di qualità" sulla decisione a garanzia di potenziali errori del sistema decisionale automatizzato: il patrimonio conoscitivo dell'essere umano, le intuizioni, le capacità di effettuare bilanciamenti tra i diritti, possono essere utili per identificare errori che produrrebbero danni su larga scala[112]. Vi sono fattori da valutare che non possono essere tradotti in linguaggio matematico e soppesati adeguatamente da un algoritmo: per questo è necessario garantire che i soggetti non siano ridotti a meri fatti matematicamente computabili[113].

Ulteriori caratteri della decisione sono l'*effetto legal*e che deve produrre, cioè l'idoneità a pregiudicare la sfera giuridica del soggetto e incidere *significativamente* sulla persona, cioè sulle sue scelte e i suoi comportamenti.

Di fatto, talune decisioni che comportano l'esclusione da alcune opportunità (si pensi al rifiuto del credito, alle procedure di assunzione) incidono in maniera significativa sulla persona considerata, di conseguenza il Regolamento ha voluto assicurare una componente umana significativa in grado di valutare i singoli casi concreti minimizzando il potenziale discriminatorio e iniquo dei processi algoritmici, che non sono certamente in grado di effettuare un bilanciamento tra i diritti in gioco e di soppesare adeguatamente tutti i fattori coinvolti [114]. Nel momento in cui deleghiamo a un algoritmo la decisione di concedere un mutuo, deleghiamo anche i relativi aspetti etico-sociali e vi è il rischio che l'algoritmo non

[110] A. Ziroldi, *Intelligenza artificiale e processo penale tra norme, prassi e prospettive*, in *Questione di Giustizia,*2019
[111] *Infra* par. 3
[112] M. Almada,, *Human Intervention in Automated Decision-Making: Toward the Construction of Contestable Systems* ,2019, Forthcoming, 17th International Conference on Artificial Intelligence and Law (ICAIL 2019)
[113] *Ibidem*
[114] C. Serra,*Il diritto di contestazione delle decisioni automatizzate nel GDPR* in *Anuario de la Facultad de derecho de la Universidad de Alcalà,* Vol. XII, 2019

consideri ragioni pertinenti e non discriminatorie: vi è il rischio che sia negato a soggetti potenzialmente idonei ma ritenuti inaffidabili perchè provenienti ad esempio da un quartiere dove risiede un'alta percentuale di soggetti con comportamenti finanziari inaffidabili[115].

Al paragrafo 2 dell'art. 22, il Regolamento precisa che "*Il paragrafo 1 non si applica nel caso in cui la decisione:*

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[116], 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."

In ogni caso la praticabilità delle eccezioni è subordinata all'adozione da parte del titolare del trattamento di adeguate misure di tutela dei diritti, libertà e interessi della persona coinvolta. Il paragrafo 3 dell'art. 22 prevede che sia garantito "*almeno il diritto di ottenere l'intervento umano da parte del titolare del trattamento, di esprimere la propria opinione e di contestare la decisione".* La specificazione manca invece in relazione alla fattispecie b), probabilmente per assicurare all'Unione o allo Stato membro una maggiore libertà di determinazione delle misure adeguate di tutela da adottare[117].

Viene dunque sancito il *diritto di contestazione* della decisione basata unicamente su un processo automatizzato per i casi enunciati alle lettere a) e c). Per completezza espositiva, è sufficiente sottolineare la distinzione tra il diritto di contestazione alla *decisione* contenuto nell'art. 22 comma 3 dal diritto di opposizione al *trattamento* contenuto all'art. 21. Mentre quest'ultimo agisce come

[115] S. Quintarelli*, ult.op.cit.*, p.79

[116] Il titolare del trattamento può essere autorizzato ad impiegare la decisione automatizzata anche se la decisione è diretta verso individui soggetti ad altri Stati membri.

[117] *Supra* nota 113

un veto al trattamento stesso dei dati, intervenendo a monte, il diritto di contestazione interviene sull'esito del trattamento stesso, quando viene esplicata la decisione, in qualità di atto di difesa.[118] Infatti, imporre al titolare del trattamento di garantire il diritto di contestazione significa garantire forme di gestione della controversia rispettose dei diritti processuali, quali il diritto al contraddittorio, il diritto alla prova e il diritto ad una decisione equidistante[119]. Possiamo leggere dunque in tale articolo la previsione di rimedi dalla tutela progressiva: dal *minimum* dell'intervento umano alla possibilità di contestare la decisione.

2.2.2 La Direttiva UE 680/2016 in materia di trattamento dei dati personali ai fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali

La disciplina europea in materia di trattamento automatizzato dei dati personali contenuta nel GDPR si completa con la Direttiva UE 680-2016[120], adottata il 27 aprile 2016 dal Parlamento e dal Consiglio europeo in materia di trattamento interamente o parzialmente automatizzato dei dati personali da parte delle autorità competenti a fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali. Il trattamento dei dati in questo ambito può essere svolto solo da una autorità competente.[121]

L'Art. 11 comma 1 riproduce il contenuto dell'Art. 22 GPDR in materia di decisione basata unicamente su un trattamento automatizzato, statuendo che "una decisione basata unicamente su un trattamento automatizzato, compresa la

[118] *Ibidem*
[119] *Ibidem*
[120] Tale Direttiva è stata recepita nel nostro ordinamento con il Decreto Legislativo 21.5.2018 n. 51, la cui portata per la materia penale sarà analizzata al par. 3 .
[121] «Qualsiasi autorità pubblica dello Stato, di uno Stato membro dell'Unione europea o di uno Stato terzo competente in materia di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali, incluse la salvaguardia contro e la prevenzione di minacce alla sicurezza pubblica; 2) qualsiasi altro organismo o entità incaricato dagli ordinamenti interni di esercitare l'autorità pubblica e i poteri pubblici a fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali, incluse la salvaguardia e la prevenzione di minacce alla sicurezza pubblica.»

profilazione, che produca effetti giuridici negativi o incida significativamente sull'interessato sia vietata, salvo che sia autorizzata dal diritto dell'Unione o dello Stato membro cui è soggetto il titolare del trattamento e che preveda garanzie adeguate per i diritti e le libertà dell'interessato" tra cui almeno il diritto di ottenere l'intervento umano da parte del titolare del trattamento.

La previsione è rafforzata dal Considerando 38, in virtù del quale l'interessato "dovrebbe avere il diritto di non essere oggetto di una decisione che valuta aspetti personali che lo concernono basata *esclusivamente* su un trattamento automatizzato e che produca effetti giuridici negativi nei suoi confronti o incida significativamente sulla sua persona"; devono essere in ogni caso fornite garanzie adeguate tra cui il diritto a ottenere l'intervento umano, il diritto di esprimere la propria opinione, di ottenere una spiegazione della decisione raggiunta dopo tale valutazione e di impugnare la decisione.

Inoltre, si prevede che la discriminazione di persone fisiche sulla base di dati personali che, per loro natura, sono particolarmente sensibili sotto il profilo dei diritti e delle libertà fondamentali, dovrebbe essere vietata.

Infine, l'art. 29 comma 2 dispone che il titolare del trattamento attui una serie di misure volte ad impedire che persone non autorizzate utilizzino sistemi di trattamento automatizzato mediante attrezzature per la trasmissione di dati (lett. d),

garantire che le persone autorizzate a usare un sistema di trattamento automatizzato abbiano accesso solo ai dati personali cui si riferisce la loro autorizzazione d'accesso (lett. e)

e garantire la possibilità di verificare e accertare a posteriori quali dati personali sono stati introdotti nei sistemi di trattamento automatizzato, il momento della loro introduzione e la persona che l'ha effettuata («controllo dell'introduzione») (lett. g).

2.2.3 La Carta etica europea sull'uso dell'intelligenza artificiale nei sistemi giudiziari

Rispetto alla realtà statunitense, l'Europa è parsa molto più cauta nell'introdurre le conquiste della "quarta rivoluzione" all'interno della macchina giudiziaria. La Carta Etica assume un significato particolare proprio perché segna la presa di coscienza da parte dell'Europa delle enormi potenzialità dell'AI e la possibilità di sfruttarla anche nella realtà giudiziaria, se declinata in un certo modo[122].

La Commissione europea per l'efficacia della giustizia (CEPEJ[123]) del Consiglio d'Europa ha adottato nel Dicembre 2018 la Carta etica europea sull'uso dell'intelligenza artificiale nei sistemi giudiziari, fissando i principi etici da osservare per l'impiego dell'AI nei sistemi giudiziari.

La Carta è destinata ad attori[124] pubblici e privati che sviluppano servizi e strumenti di AI in tale ambito[125] ed è corredata da quattro Appendici che presentano una panoramica dello stato dell'arte nei sistemi giudiziari europei.

Nell'introduzione viene riconosciuto il potenziale applicativo degli strumenti di IA per implementare l'efficienza del sistema giudiziario (che contribuirebbero a migliorare la prevedibilità della legge e la coerenza delle decisioni giudiziarie) ma viene anche evidenziata l'importanza di un suo uso nel rispetto dei diritti umani, in particolare di quelli sanciti nella CEDU e della Convenzione sulla protezione delle persone rispetto al trattamento automatizzato di dati di carattere personale.

All'interno della Carta vengono enunciati cinque principi:

I. Il principio del rispetto dei diritti fondamentali

II. Il principio di non-discriminazione

---

[122] S. Quattrocolo*, Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale "predittiva"* in *Cassazione penale* n .4/2019, p. 203

[123] Istituita nel 2002 per iniziativa del Comitato dei Ministri del Consiglio d'Europa per monitorare e misurare la qualità dei sistemi giudiziari dei Paesi membri.

[124] Sia operatori del diritto sia sviluppatori di strumenti della IA (imprese private, soggetti pubblici e autorità coinvolte nello sviluppo di tali tecnologie in materia di servizi legali).

[125] CEPEJ, *Carta Etica etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi* adottata dalla CEPEJ nel corso della sua 31ª Riunione plenaria (Strasburgo, 3-4 dicembre 2018)

III.Il principio di qualità e sicurezza

IV. Il principio di trasparenza, imparzialità e equità

V. Il principio del controllo da parte dell'utilizzatore

In virtù del primo principio, il trattamento di decisioni e dati giudiziari deve avere finalità chiare, che rispettino pienamente i 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. Quando utilizzati in ambito giudiziario, tali strumenti non devono ledere il diritto a un equo processo (comprensivo del principio della parità delle armi e del contraddittorio), a un giudice terzo e imparziale ed il diritto di accesso al giudice[126].

Il principio di non discriminazione è finalizzato a prevenire "lo sviluppo o l'intensificazione di qualsiasi discriminazione tra persone o gruppi di persone", assicurando metodologie che non riproducano o aggravino dinamiche discriminatorie tra i gruppi sociali che potrebbero verificarsi in virtù della raccolta e classificazione di dati (e che possono comprendere origine etnica, fede religiosa, convinzioni politiche, condizioni socio-economiche ecc.), con particolare riferimento ai *Risk assessment tools*[127] impiegati soprattutto in USA e UK. I soggetti potrebbero essere vittime di *implicit bias* sia nel caso in cui l'input non sia completamente neutro, sia nel caso in cui l'algoritmo riproduca preconcetti sociali[128].

Il principio di qualità e sicurezza dei dati da utilizzare raccomanda "l'uso di fonti certificate e dati intangibili, attraverso modelli concepiti in modo multidisciplinare, in un ambiente tecnologico sicuro". È necessario impiegare fonti certificate e

[126] *Ibidem*

[127] Si pensi all'impiego degli algoritmi predittivi in materia penale come COMPAS, usato negli USA per calcolare il rischio di recidiva degli accusati nel processo. L'algoritmo prende in considerazione numerosi fattori, tra cui l'origine etnica e le condizioni socio-economiche dei soggetti sottoposti alla decisione; uno studio condotto da ProPublica ha evidenziato come gli afroamericani abbiano il doppio della probabilità di essere considerati recidivi rispetto a altri gruppi etnici. La questione sarà analizzata nel Capitolo III, par. 7

[128] S. Quattrocolo, *Intelligenza artificiale e giustizia: nella cornice della Carta Etica europea, gli spunti per un'urgente discussione tra scienze penali e informatiche* in *La Legislazione Penale*,2018

garantire la completezza e integrità dei dati impiegati, predisponendo modelli funzionali di *machine learning* multidisciplinari che integrino competenze di professionisti del settore (giudici, pubblici ministeri, avvocati, docenti nei campi del diritto). Di fatto, la scelta dei dati da impiegare implica un'attenta analisi dell'integrità della fonte e del dato stesso per evitare che siano modificati. L'intero processo deve pertanto essere tracciabile, al fine di garantire che non abbia avuto luogo alcuna modifica in grado di alterare il contenuto o il significato della decisione trattata. Per gli stessi motivi, i modelli e gli algoritmi su cui si fonda l'elaborazione devono essere custoditi in ambienti sicuri.

Il principio di trasparenza, imparzialità e equità raccomando "l'accessibilità, la comprensibilità e la verificabilità esterna dei processi computazionali" fissando il *right to explanation* del risultato decisionale (inteso come comprensibilità dello stesso) e il diritto di accesso al meccanismo di funzionamento interno dell'algoritmo, spesso ostacolato dal segreto industriale, coinvolgendo esperti esterni nella certificazione di tali qualità; dovrebbero essere dunque resi noti quantomeno informazioni parziali fondamentali in materia di algoritmi, per esempio quali siano le variabili utilizzate, quali siano gli obiettivi cui è finalizzata l'ottimizzazione degli algoritmi, i dati di apprendimento, i valori medi e gli scarti tipo dei risultati ottenuti, o la quantità e il tipo di dati trattati dall'algoritmo[129].

Il quinto principio prevede un uso degli algoritmi "garantendo che gli utilizzatori agiscano come soggetti informati, nel pieno controllo delle loro scelte". L'utilizzatore del sistema deve avere la possibilità di rivedere in qualsiasi momento la decisione giudiziaria e i dati utilizzati per ottenere il risultato senza essere vincolati ad esso.

Nell'Appendice II della carta si esaminano gli utilizzi dell' AI nei sistemi giudiziari europei distinguendo tra *utilizzi che devono essere incoraggiati* (comprendenti lo sviluppo di motori di ricerca giurisprudenziali avanzati, tecniche di apprendimento automatico per migliorare le banche dati esistenti, chatbot per facilitare l'accesso alle informazioni esistenti), *usi possibili che esigono notevoli*

[129] Proposta effettuata a pagina 38 dello studio della MSI-NET del Consiglio d'Europa in materia di *"Algoritmi e diritti umani"*

precauzioni metodologiche (ad esempio per quanto riguarda il supporto a misure alternative di risoluzione delle controversie in materia civile, l'impiego di strumenti di polizia predittivi per individuare luoghi in cui saranno commessi i reati, piattaforme di risoluzione online di *small claims*), *utilizzi da esaminare al termine di supplementari studi scientifici* (per la profilazione dei magistrati e l'anticipazione delle decisioni dei tribunali) e *utilizzi da esaminare con le più estreme riserve.* In particolare, quest'ultima categoria si riferisce all'utilizzo di strumenti di valutazione del rischio in materia penale, come l'algoritmo COMPAS[130] negli Stati Uniti o HART [131] nel Regno Unito, che sfruttando un approccio meramente statistico conducono spesso a risultati discriminatori o errati[132]. La Carta deve essere intesa come "uno spazio in cui sono stati tracciati dei confini": saranno poi gli attori a delineare politiche di accountability, strategie, meccanismi di funzionamento e tutele adeguate, attraverso un sistema di *checks and balances.*[133]

[130] Per il calcolo del rischio di recidiva in ambito processuale penale. In argomento i*nfra* capitolo IV, par.7
[131] Utilizzato per calcolare il rischio di recidiva dell'indiziato e in ambito di polizia, per prevedere dove avverano i crimini con più probabilità. *Infra c*apitolo IV, par. 6.1
[132] Sul punto vedere l'Appendice II della *Carta Etica*
[133] Per una riflessione più approfondita consultare C. Castelli, D. Piana, Giusto processo e intelligenza artificiale, Maggioli, Santarcangelo di Romagna, 2019, pp. 109 et seq.

CAPITOLO III

ORDINAMENTO ITALIANO ED INTELLIGENZA ARTIFICIALE

3.1 Algoritmi e ordinamento italiano: i recenti sviluppi giurisprudenziali in tema di decisione automatizzata

Come si dimostrerà nel corso di questa trattazione[134], il campo del diritto non è stato immune dalla penetrazione dell'intelligenza artificiale. Negli ordinamenti nazionali si fa sempre più ricorso agli strumenti di automazione sviluppati con l'avvento della "quarta rivoluzione", fino ad essere impiegati sempre più frequentemente nei processi civili, amministrativi e finanche penali (come accade negli Stati Uniti con l'utilizzo degli algoritmi predittivi).

L'uso degli algoritmi nel processo penale è ancora un'utopia nel nostro ordinamento (in virtù delle garanzie individuate dall'art. 111 co. 2 della Costituzione[135] in materia di equo processo e dall'art. 220[136] cpp in materia di perizia), tuttavia di recente la giurisprudenza ha ammesso un uso limitato di tali strumenti nel processo amministrativo, nonostante le posizioni del Consiglio di Stato e dei tribunali regionali non siano del tutto allineate sul punto[137].

Al di là del timido riferimento alla possibilità di impiegare strumenti telematici in ambito amministrativo, contenuto nell' art. 3-bis della L. 241/90, non si rinviene alcun esplicito riferimento alla possibilità di impiegare algoritmi all'interno delle procedure amministrative: è stata dunque la giurisprudenza, sensibile agli indirizzi

---

[134] *Infra* capitolo III par. 2

[135] «Ogni processo si svolge nel contraddittorio tra le parti, in condizioni di parità, davanti a giudice terzo e imparziale. La legge ne assicura la ragionevole durata.»

[136] «Non sono ammesse perizie per stabilire l'abitualità o la professionalità nel reato, la tendenza a delinquere, il carattere e la personalità dell'imputato e in genere le qualità psichiche indipendenti da cause patologiche».

[137] I primi sono tendenzialmente restii ad ammettere l'uso di questi strumenti automatizzati, mentre il Consiglio di stato, di recente, ha aperto a un impiego di tali strumenti seppur limitato e nel rispetto delle garanzie e principi fondamentali enunciati a livello europeo.

provenienti dall'ordinamento comunitario, a fornire i primi arresti e i primi segnali di apertura nei confronti degli strumenti di automazione [138]. La principale preoccupazione derivante dall'utilizzo degli algoritmi è il loro meccanismo di funzionamento, basato sul *Machine Learning*: non è sempre possibile, infatti, stabilire in che modo l'algoritmo sia giunto a quello specifico risultato, né quali siano stati i fattori coinvolti nella decisione e il peso loro attribuito. Ciò appare ancor più problematico in ambito amministrativo poiché, alla luce dei principi di buon andamento della pubblica amministrazione, la stessa deve essere sempre in grado di motivare e spiegare adeguatamente le sue decisioni e il procedimento alla base.

3.1.1 L'impossibilità di sostituire l'attività valutativa umana con un algoritmo:la pronuncia n. 9224/2018 del Tar del Lazio

A seguito del piano straordinario di assunzioni a tempo indeterminato e mobilità su scala nazionale avviato nel 2015 dal M.I.U.R, sono state promosse assunzioni di docenti attraverso un piano di trasferimenti interprovinciale del personale (c.d. mobilità della "buona scuola").

Per velocizzare e semplificare l'attività dell'amministrazione per l'assegnazione della sede spettante al singolo docente è stato utilizzato un algoritmo; i docenti hanno impugnato la graduatoria di mobilità (che prevedeva il trasferimento in province più lontane da quella della propria residenza o da quella indicata), lamentando che si fosse "demandato a un impersonale algoritmo lo svolgimento dell'intera procedura di assegnazione dei docenti[139]" e che l'algoritmo

[138] D.U. Galetta, J. G. Corvalán, *Intelligenza Artificiale per una Pubblica Amministrazione 4.0? Potenzialità, rischi e sfide della rivoluzione tecnologica in atto*, in Federalismi, n.3/ 2019.

[139]Tar Lazio, sez. III *bis*, nn. 9224-9230 del 10 settembre 2018; in particolare «i ricorrenti denunciano che il delineato piano straordinario non è stato corredato da alcuna attività amministrativa ma è stato demandato ad un algoritmo, tuttora sconosciuto, per effetto del quale sono stati operati i trasferimenti e le assegnazioni in evidente contrasto con il fondamentale principio della strumentalità del ricorso all'informatica nelle procedure amministrative».

avesse "completamente sostituito l'istruttoria commessa ad un ufficio e ad un responsabile"[140].

Il Tar del Lazio, accogliendo i ricorsi, ha censurato l'assenza di una effettiva attività amministrativa alla base delle assegnazioni: l'algoritmo è infatti del tutto "impersonale e orfano di capacità valutazionali delle singole fattispecie concrete tipiche della garantistica istruttoria procedimentale[141]" e determina un *vulunus* non solo per il diritto di difesa e partecipazione al procedimento, ma anche per l'obbligo di motivazione delle decisioni amministrative di cui all'art. 24 della Costituzione (non potendosi ricostruire l'*iter* logico alla base della decisione). Il Tar conclude sul punto che "gli istituti di partecipazione, di trasparenza e di accesso, in sintesi, di relazione del privato con i pubblici poteri non possono essere legittimamente mortificate e compresse soppiantando l'attività umana con quella impersonale".

Il Tribunale ha inoltre precisato che gli algoritmi "finanche ove pervengano al loro maggior grado di precisione e addirittura alla perfezione, non possano mai soppiantare, sostituendola appieno, l'attività cognitiva, acquisitiva e di giudizio che solo un'istruttoria affidata ad un funzionario persona fisica è in grado di svolgere e che pertanto, (…) deve seguitare ad essere il *dominus* del procedimento stesso, all'uopo dominando le stesse procedure informatiche predisposte in funzione servente e alle quali va dunque riservato tutt'oggi un ruolo strumentale e meramente ausiliario in seno al procedimento amministrativo e giammai dominante o surrogatorio dell'attività dell'uomo".

Dalla pronuncia si evince pertanto che gli algoritmi dovrebbero mantenere una posizione meramente servente e ausiliare rispetto all'attività del funzionario amministrativo, certamente non sostituibile da un sistema di Intelligenza Artificiale nelle sue competenze, conoscenze e responsabilità. Solo in tal modo è possibile evitare la lesione delle garanzie processuali costituzionalmente garantite e delle posizioni giuridiche soggettive dei singoli soggetti coinvolti[142].

[140] Cfr. Tar Lazio, sez. III *bis*, nn. 9224 del 10 settembre 2018

[141] *Supra* nota 139, par. 3.1 della sentenza.
[142] *Supra* nota 139, par. 4.1 della sentenza

Come sarà esposto nei successivi paragrafi, in una successiva pronuncia il Consiglio di Stato ha assunto una posizione differente da quella del Tar, incoraggiando l'utilizzo degli strumenti di automazione nel procedimento amministrativo, ma subordinandone l'uso al rispetto di determinate condizioni di legittimità.

3.1.2 Il diritto di accesso all'algoritmo nelle procedure valutative della pubblica amministrazione: la sentenza 8 aprile 2019, n. 2270 del Consiglio di Stato

La sentenza in esame assume rilievo per le considerazioni logico-giuridiche alla base, che individuano le condizioni di legittimità a cui subordinare l'uso degli algoritmi nelle procedure valutative della pubblica amministrazione. Rispetto alla precedente sentenza del Tar, il Consiglio di Stato mostra una maggiore apertura nei confronti della possibilità di utilizzare le decisioni automatizzate, incoraggiandone addirittura l'utilizzo in virtù degli indiscutibili vantaggi derivanti dalla automazione del processo. Tali vantaggi sono apprezzabili soprattutto con riferimento a procedure seriali o standardizzate, implicanti l'elaborazione di ingenti quantità di istanze prive di ogni apprezzamento discrezionale[143].

Nella sentenza si afferma che l'utilizzo di una procedura informatica che conduca direttamente alla decisione finale non deve essere stigmatizzata, ma anzi incoraggiata. Essa comporta, infatti, numerosi vantaggi quali, ad esempio, la notevole riduzione della tempistica procedimentale per operazioni meramente ripetitive e prive di discrezionalità, risultando coerente declinazione dell'art. 97 Cost. L'utilizzo di dette procedure, tuttavia, non può sostanziarsi in una elusione dei princìpi che conformano il nostro ordinamento e che regolano lo svolgersi dell'attività amministrativa. La regola tecnica che governa ciascun algoritmo "resta pur sempre una *regola amministrativa* generale, costruita dall'uomo e non dalla macchina"[144]. Si precisa inoltre che l'algoritmo deve essere considerato a tutti gli

[144] Sentenza del Consiglio di Stato, sez. VI, 8 aprile 2019, n. 2270

effetti un *atto amministrativo informatico:* ciò implica necessariamente il rispetto di una serie di principi fondamentali individuati anche a livello comunitario, come il *principio di trasparenza algoritmica,* in virtù del quale il meccanismo tramite il quale si concretizza la decisione robotizzata deve essere sempre conoscibile.

Tale conoscibilità deve essere garantita in tutti gli aspetti: dai suoi autori al procedimento usato per la sua elaborazione, al meccanismo di decisione, comprensivo delle priorità assegnate nella procedura valutativa e decisionale e dei dati selezionati come rilevanti[145].

In secondo luogo, la regola algoritmica deve essere non solo conoscibile in sé, ma anche sindacabile dal giudice amministrativo: è necessario garantire l'imputabilità della decisione all'organo titolare del potere (*principio di responsabilità),* il quale deve poter effettuare una verifica in termini di logicità e di correttezza degli esiti decisionali dell'algoritmo[146]. Alla luce di tali considerazioni è stato ritenuto fondato l'appello dei docenti che contestavano l'illogicità del meccanismo di assegnazione alla sede di servizio non richiesta, effettuato mediante un algoritmo di cui non si conosceva il metodo di funzionamento, risultando soccombenti rispetto a colleghi con punteggio inferiore in graduatoria. Il Consiglio di Stato ha infatti accertato la violazione dei principi di di imparzialità, pubblicità e trasparenza "poiché non è dato comprendere per quale ragione le legittime aspettative di soggetti collocati in una determinata posizione in graduatoria siano andate deluse". Proseguendo sul punto, si statuisce che "l'impossibilità di comprendere le modalità con le quali, attraverso il citato algoritmo, siano stati assegnati i posti disponibili, costituisce di per sé un vizio tale da inficiare la procedura."

---

[145] Ciò «al fine di poter verificare che gli esiti del procedimento robotizzato siano conformi alle prescrizioni e alle finalità stabilite dalla legge o dalla stessa amministrazione a monte di tale procedimento e affinché siano chiare – e conseguentemente sindacabili – le modalità e le regole in base alle quali esso è stato impostato».

[146] «La suddetta esigenza risponde infatti all'irrinunciabile necessità di poter sindacare come il potere sia stato concretamente esercitato, ponendosi in ultima analisi come declinazione diretta del diritto di difesa del cittadino, al quale non può essere precluso di conoscere le modalità (anche se automatizzate) con le quali è stata in concreto assunta una decisione destinata a ripercuotersi sulla sua sfera giuridica».

Sulla stessa questione è intervenuta la successiva pronuncia n. 8472/2019[147] del Consiglio di Stato, relativa ad un analogo caso di mobilità interprovinciale dei docenti. Il giudice amministrativo, dopo aver ribadito il principio per cui la "formula tecnica", cioè l'algoritmo, deve essere traducibile nella "regola giuridica" ad essa sottesa in modo tale da essere leggibile e comprensibile[148], statuisce che "*non può assumere rilievo l'invocata riservatezza delle imprese produttrici dei meccanismi informatici utilizzati i quali, ponendo al servizio del potere autoritativo tali strumenti, all'evidenza ne accettano le relative conseguenze in termini di necessaria trasparenza*"[149]. Viene dunque riaffermato il diritto di accesso al codice sorgente dell'algoritmo, funzionale alla comprensione del funzionamento del *software* e si pone un significativo ostacolo al problema dell'opacità dell'algoritmo.

3.1.3 Il trattamento automatizzato dei dati personali per finalità di prevenzione e repressione dei reati alla luce del decreto legislativo n. 51/2018

In adeguamento alla normativa europea in materia di trattamento dei dati personali, è stato recepito all'interno del nostro ordinamento con il decreto legislativo n. 51 del 18 Maggio 2018 in attuazione alla direttiva UE 680/2016[150] relativa alla protezione delle persone fisiche con riguardo al trattamento dei dati personali da parte delle autorità competenti a fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali.

Il decreto legislativo, che disciplina il trattamento[151], interamente o parzialmente automatizzato, dei dati personali per finalità di prevenzione e repressione dei reati,

[147] Cons. Stato Sez. VI, Sent., 13-12-2019, n. 8472
[148] G. Pesce, *ult. op. cit.*
[149] *Supra* nota 146, par. 13.1. della motivazione.

[151] Nello specifico sono descritte le modalità di trattamento dei dati e di conservazione, nonché i diritti del soggetto sottoposto al trattamento e gli obblighi da osservare per assicurare un trattamento non pregiudizievole degli stessi; è prevista inoltre l'istituzione di

esecuzione di sanzioni penali, salvaguardia e prevenzione di minacce alla sicurezza pubblica, è particolarmente rilevante poiché estende anche in sede penale le cautele contenute all'interno del GDPR. In particolare, l'art. 8 del decreto, derubricato "processo decisionale automatizzato relativo alle persone fisiche", riproduce quasi integralmente il contenuto dell'Art. 22 del GPDR ma, a differenza di questo, sancisce il divieto assoluto di decisioni basate unicamente su un trattamento automatizzato in ambito penale. Si legge infatti: "1. Sono vietate le decisioni basate unicamente su un trattamento automatizzato, compresa la profilazione, che producono effetti negativi nei confronti dell'interessato, salvo che siano autorizzate dal diritto dell'Unione europea o da specifiche disposizioni di legge. 2. Le disposizioni di legge devono prevedere garanzie adeguate per i diritti e le libertà dell'interessato. In ogni caso è garantito il diritto di ottenere l'intervento umano da parte del titolare del trattamento."

È infine specificato che le decisioni di cui al punto uno non possono basarsi sulle categorie particolari di dati personali (relativi all'appartenenza a razza, religione, sesso etc.), salvo che siano in vigore misure adeguate a salvaguardia dei diritti, delle libertà e degli interessi legittimi dell'interessato.

Lo stesso decreto ha previsto inoltre due nuovi illeciti penali: il reato di profilazione finalizzata alla discriminazione e quello di trattamento illecito di dati sensibili.

3.2 Profili costituzionali e penalistici di rilievo

Ai fini di una migliore contezza del fenomeno, si cercherà di condurre un ragionamento sul rischio di ingresso di algoritmi predittivi nel processo penale nazionale, analizzando gli sbarramenti normativi che ne impedirebbero l'infiltrazione. *In primis* si consideri che nel nostro sistema penale la commisurazione della pena è affidata alla valutazione *discrezionale* del giudice:

un responsabile per la protezione dei dati (nello specifico il Garante per la protezione dei dati personali).

affidare il giudizio prognostico in termini di gravità del reato e capacità di delinquere (art. 133 c.p) collliderebbe con alcuni basilari principi dell'ordinamento[152].

3.2.1 I limiti costituzionali

Rispetto al sistema penale americano, nell'ordinamento penale italiano non esiste una cesura temporale tra pronuncia della sentenza e irrogazione della pena, tantomeno una fase di istruttoria sulla personalità del reo[153].

L'ipotesi che un algoritmo possa sostituire o anche semplicemente affiancare il giudice nella valutazione del rischio di un reo è preclusa da una serie di "paletti" costituzionali: oltre ai limiti contenuti negli artt. 25 («nessuno può essere distolto dal giudice naturale precostituito per legge») e 102 («la funzione giurisdizionale è esercitata da magistrati ordinari istituiti e regolati dalle norme sull'ordinamento giudiziario»), l'art. 101, comma 1, Cost., nel disporre che i giudici sono soggetti soltanto alla legge, esclude che il giudice possa essere vincolato dall'esito di procedure algoritmiche[154].

Un limite fondamentale all'introduzione degli algoritmi è garantito dall'art. 111, comma 4, Cost. che garantisce il contraddittorio nella formazione della prova, impedendo al giudice di acquisire o di valutare elementi diversi da quelli oggetto di contraddittorio tra le parti[155].

---

[152] L. D'agostino, *Gli algoritmi predittivi per la commisurazione della pena* in *Diritto Penale Contemporaneo,* 2/2019, pp. 267 e ss.
[153] *Ibidem*, p. 366
[154] F. Donati, *Intelligenza artificiale e giustizia* in *Rivista Associazione italiana dei Costituzionalisti* N°: 1/2020, p. 428
[155] *Ibidem*

3.2.2 Il giudizio di pericolosità sociale nella normativa italiana alla luce dell'art 203 c.p.

Ai sensi dell'art. 203[156], co.1 c.p., la pericolosità sociale può essere definita come la *probabilità* che il soggetto in futuro commetta nuovi reati e il suo accertamento è condizione per l'erogazione di misure di sicurezza personale.

A seguito degli interventi della Corte Costituzionale[157], non sono più ammissibili presunzioni di pericolosità: la pericolosità sociale deve essere sempre accertata *in concreto* dal giudice[158]. Tale giudizio si articola in due momenti: quello dell*'analisi della personalità* del soggetto e della *prognosi criminale* conseguente.

La qualità di «persona socialmente pericolosa» si desume dalle circostanze indicate nell'art. 133 c.p: bisogna tener conto di:

- *Gravità del reato* desunta da:
    1. dalla natura, dalla specie, dai mezzi, dall'oggetto, dal tempo, dal luogo e da ogni altra modalità dell'azione;
    2. dalla gravità del danno o del pericolo cagionato alla persona offesa dal reato;
    3. dall'intensità del dolo o dal grado della colpa.

- *Capacità a delinquere* desunta da:
    1. dai motivi a delinquere e dal carattere del reo;
    2. dai precedenti penali e giudiziari e, in genere, dalla condotta e dalla vita del reo, antecedenti al reato;
    3. dalla condotta contemporanea o susseguente al reato;
    4. dalle condizioni di vita individuale, familiare e sociale del reo.

[156] «Agli effetti della legge penale, è socialmente pericolosa la persona, anche se non imputabile o non punibile, la quale ha commesso taluno dei fatti indicati nell'articolo precedente, quando è probabile che commetta nuovi fatti preveduti dalla legge come reati.»
[157] Corte cost. 27 luglio 1982, n. 139; Corte cost. 28 luglio, n.249
[158] G.Marinucci, E. Dolcini. G.L Gatta, *Manuale di diritto penale- Parte Generale,* Giuffrè , Milano , 2019, pp. 807 e ss.

Il giudizio di capacità a delinque si proietta necessariamente nel futuro ai fini di una prognosi sulle probabilità di commissione di un nuovo reato o di un reato dello stesso tipo, ma non ha valore esplicitamente predittivo.

La valutazione del carattere del reo richiede una complessa valutazione della sua personalità e delle caratteristiche *innate* del soggetto idonee a orientare i suoi comportamenti (es. capacità di autocontrollo e stabilità emotiva)[159]. Sono oggetto di valutazione anche le condizioni sociali e familiari del soggetto (emarginazione, disoccupazione, adesione a bande criminali ecc.)

Proprio per la complessità dei fattori da valutare è stabilito che il giudice effettui l'esame autonomamente, sulla base di un ragionamento razionale e logico che tenga conto delle fragilità del reo, senza poter ricorrere a valutazioni tecnico-scientifiche. Di fatto l'algoritmo fornisce una maggiore certezza e oggettività rispetto alle valutazioni umane solo *apparentemente*[160]. Come si dimostrerà nei successivi capitoli, la qualità dei dati inseriti e le correlazioni statistiche "viziate" da *bias* possono fornire un risultato non aderente alla realtà. Come osservato «*even with masses of data, there is no automathic tecnhique for turning correlation into causation*[161].»

3.2.3 L' Art. 220 c.p.p. e il divieto di perizie per stabilire la tendenza a delinquere

L'art 220 c.p.p. è considerato il "baluardo" a difesa dell'ingresso degli algoritmi nel processo, proprio perchè afferma l'inammissibilità di perizie per «stabilire l'abitualità o la professionalità nel reato, la tendenza a delinquere, il carattere e la personalità dell'imputato e in genere le qualità psichiche indipendenti da cause patologiche».

---

[159] *Ibidem*
[160] V. Manes, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in *Discrimen*, 2020.
[161] D.J. Spiegelhalther*, The Future lies in Uncertainty*, in *Science*, 2014, vol. 435, 264

La perizia sarebbe finalizzata a delineare un profilo della personalità e del carattere del reo per individuare la pena o la misura di sicurezza più adatta al caso di specie.

Il legislatore si è fortemente opposto al ricorso a tale tecnica "automatizzata" di analisi: la *ratio* del divieto si può rinvenire nella necessità di tutelare la libertà morale dell'imputato, giacché si delineerebbe il rischio di cedere ai pregiudizi inerenti a aspetti particolari del carattere dell'imputato che potrebbero condizionare l'organo giudicante.

Si vuole dunque evitare che il giudice, nell'assumere le sue determinazioni, si basi essenzialmente sull' *identità dell'imputato* tracciata dalle perizie psicologiche e non sui *fatti commessi* in concreto[162].

Corollario logico di tale previsione è il divieto di servirsi di strumenti algoritmici di valutazione del rischio (come COMPAS[163]), sostanzialmente equiparabili a perizie criminologiche. Il legislatore dimostra così di rifiutare in radice l'idea di un'istruttoria sulla capacità a delinquere del reo[164] (a differenza dei sistemi d'oltreoceano incentrati su tale modello), introducendo una ritrosia aprioristica circa l'attendibilità della prova scientifica in materia di personalità dell'imputato. Tale ritrosia è connessa alla labilità dell'indagine e del rischio di violazione del diritto di difesa, laddove potrebbero facilmente essere aggirate le garanzie e gli strumenti tipici per l'acquisizione della prova[165].

> La prognosi sulla capacità a delinquere è un giudizio *intuitu personae* e nessuna valutazione statistica può supportare o sostituire un giudizio di questo tipo, anche perchè «il punteggio di rischio verrebbe in tal modo calcolato incrociando i dati relativi a situazioni simili o vicende analoghe, facendo cadere il giudizio sulla pericolosità sociale del reo in un labirinto di inevitabili generalizzazioni empiriche[166].»

---

[162] A. Di Prisco, *Elementi di criticità sulla perizia psicologica nel processo penale*, in Ius in Itinere, 2018
[163] *Infra* Cap. IV, para.7
[164] *Supra* nota 152
[165] *Ibidem*
[166] *Ibidem*

Fondando il giudizio sulla condotta *hic et nunc* posta in essere dal soggetto, viene scongiurato il rischio che la valutazione si fondi in esclusiva su condotte poste in essere, in passato, da soggetti che si siano trovati in situazione analoghe all'imputato in esame[167].

3.3 Il problema della responsabilità da algoritmo

La rapida evoluzione degli scenari aperti dai sistemi algoritmici impongono nuove riflessioni anche in tema di "responsabilità da algoritmo" in caso di danno.

Gli ambiti in cui può profilarsi una responsabilità di questo tipo sono estremamente numerosi: si pensi alle *self-driving car*, ai sistemi robotici utilizzati in chirurgia per l'automazione della pratica clinica, fino al ricorso all'AI per commettere crimini (ad esempio manipolazioni in ambito di mercato finanziario). La complessità di tali tecnologie rende ancora più sfumati i contorni di una responsabilità penale e civile.

Per quanto riguarda la responsabilità penale, è necessario distinguere tra danni cagionati da sistemi soggetti a controllo umano e sistemi di *Machine Learning*, capaci di apprendere autonomamente e in un modo imprevedibile anche per gli sviluppatori dello stesso algoritmo. Per quanto riguarda la prima ipotesi, si pensi alle *self-driving car* o all'impiego della robotica in ambito chirurgico. In questi casi un certo coefficiente di controllo umano sulla macchina è sempre individuabile ed è variamente graduato: da sistemi in cui vi è una semi-automazione a sistemi in cui vi è un massiccio intervento dell'uomo. Maggiore è l'autonomia del sistema, maggiori sono i problemi di individuazione del centro di responsabilità per eventi lesivi. Si potrebbe individuare una *posizione di garanzia* (inquadrabile nelle posizioni di controllo) in capo al soggetto responsabile del funzionamento del sistema. In caso di danno da funzionamento difettoso, invece, si potrebbe ricondurre il danno cagionato sul terreno della responsabilità per danno da prodotto. Volendo

[167] D. Polidoro, *Tecnologie informatiche e procedimento penale: la giustizia penale "messa alla prova" dall'intelligenza artificiale* in *Archivio Penale*, 2020, n. 3

estrarre un principio in tal senso, viene in risalto, con riguardo alla responsabilità da algoritmo, la figura del *controllore umano* che opererebbe in una duplice direzione:

1. *Meccansimo di salvaguardia* per evitare danni arrecati dal malfunzionamento della macchina;
2. *Catalizzatore di responsabilità* come responsabile degli eventi dannosi.

Nel secondo caso è necessario compiere uno forzo interpretativo maggiore per arrivare a delineare una nuova disciplina di responsabilità algoritmica, partendo dalle categorie logico-giuridiche presenti [168] . Queste macchine sono ontologicamente imprevedibili, per questo il diritto penale incontra numerose difficoltà di tipizzazione in tal senso, soprattutto in termini di nesso di causalità e colpa. Se si ammette che il sistema di AI è un *mezzo* dell'uomo, allora i danni cagionati saranno sempre riconducibili al suo programmatore, anche laddove siano eziologicamente imprevedibili[169]. Se tuttavia, ivi si innesta una azione auto-appresa del sistema, si potrebbe così spezzare il nesso di causalità ed escludere la responsabilità del programmatore[170].

Pur riuscendo a ricostruire la dinamica causale, l'elemento dell'imprevedibilità degli eventi legati all'auto-apprendimento basterebbe a escludere l'elemento soggettivo della colpa, potendosi profilare una responsabilità solo in termini di responsabilità oggettiva[171].

Bisognerebbe dunque incentrare l'analisi puntando sull'inquadramento dei sistemi automatici come soggetti di diritto e centri di imputazione della responsabilità, meritevoli di tutela ma anche passibili di sanzioni[172].

Dal punto di vista della responsabilità civile e del risarcimento del danno, invece, si è registrata una maggiore apertura: la Commissione Europea adottando le "*Ethic*

[168] C. Piergallini, *Intelligenza Artificiale: da mezzo a autore del reato?* In Rivista Italiana di Diritto e Procedura Penale, n. 4 del 2020, p 1749 e ss.
[169] *Ibidem*
[170] Nell'ottica della causalità adeguata, si potrebbe sostenere che l'evento in questione sfugge dalla predeterminazione delle decisioni del sistema.
[171] *Ibidem*
[172] *Ibidem*

Guidilines for Trustworthy AI", ha sostenuto che la responsabilità civile per danno causato da algoritmo non deve subire né una limitazione del tipo e dell'entità del danno risarcibile: «*Good AI governance should include accountability mechanisms, which could be very diverse in choice depending on the goals. Mechanisms can range from monetary compensation (no-fault insurance) to fault finding, to reconciliation without monetary compensations»*

CAPITOLO IV

LE APPLICAZIONI DELL'INTELLIGENZA ARTIFICIALE IN ALTRI ORDINAMENTI. IL MODELLO STATUNITENSE

4.1 Attuali utilizzi dell'intelligenza artificiale in giudizio

Gli strumenti di intelligenza artificiale hanno fornito valide soluzioni al problema della complessità tecnica, dei tempi e dei costi delle operazioni giudiziali, riuscendo a penetrare gradualmente anche in un settore sensibile come quello del diritto. In particolare, hanno assunto in alcuni Paesi (soprattutto nelle giurisdizioni statunitensi[173]) un peso di rilievo nel giudizio del condannato per la valutazione del rischio di recidiva, mentre in altri ordinamenti ha mantenuto un ruolo strumentale all'azione del giudice.

L'impronta dell'intelligenza artificiale nel processo è rintracciabile in molteplici contesti: dalle banche dati giurisprudenziali, all'uso di *chatbots* per l'assistenza legale, dall'analisi delle clausole contrattuali sino agli strumenti di valutazione del rischio in campo penale[174], ma le potenzialità applicative sono pressoché sconfinate. Si pensi, a titolo di esempio, alle possibilità offerte dalla *giustizia predittiva*[175], in grado di prevedere l'esito della controversia ai fini di una migliore gestione delle *small claims*[176] o per la determinazione dell'indennizzo in caso di risarcimento.

---

[173] Vengono utilizzati, come si approfondirà al par. 7 del Capitolo III, gli strumenti di valutazione del rischio.

[174] CEPEJ, *Carta Etica*

[175] La giustizia predittiva ha la finalità di prevenire il possibile esito di un giudizio, fornendo agli utenti dei dati di prevedibilità circa il successo o insuccesso della causa che si intende promuovere, scoraggiando le cause "temerarie" o che non hanno possibilità di successo a livello giudiziario e, nel caso, incentivando a seguire altre strade come quella stragiudiziale (Giusto processo e intelligenza artificiale pagina 68). A titolo di esempio è stato elaborato un algoritmo dall'University College di Londra in grado di fare previsioni attendibili nel 78% dei casi circa l'esito delle controversie davanti alla CEDU.

[176] L. De Renzis, *Primi passi nel mondo della giustizia «High Tech»: La decisione in un corpo a corpo virtuale fra tecnologia e umanità*, in A. Carleo, *ult. op.cit.* p. 148

Di seguito saranno illustrati alcuni dei campi applicativi in cui l'intelligenza artificiale ha fornito contributi particolarmente innovativi.

Circoscrivendo l'analisi all'esperienza italiana, contributi interessanti in materia di determinazione degli assegni di mantenimento e costituzione di tabelle predittive sono stati forniti dal Dipartimento di statistica dell'Università di Firenze, che nel 2007 ha elaborato il modello MoCAM[177] (Calcolo dell'Assegno di Mantenimento) che produce una stima dell'assegno di mantenimento per i figli nel caso di separazione, divorzio o rottura di una unione di fatto [178].

Altro progetto collaudato in materia ha coinvolto il Tribunale di Sondrio che, in collaborazione con il Professor Gianfranco D'Aietti, ha elaborato un algoritmo in grado di individuare gli elementi relativi ai costi di mantenimento delle persone nei diversi contesti territoriali e le stime tengono conto anche delle analisi dei comportamenti sociali tenuti sulla base dei numerosi precedenti in tema di separazione, correlati da dati economici. In tal modo è possibile elaborare parametri oggettivi e costruire tabelle obiettive di ausilio all'attività di stima del magistrato[179].

Anche nell'ambito della valutazione delle prove esistono numerosi programmi di intelligenza artificiale capaci di guidare l'apprezzamento del giudice: sono stati sviluppati diversi programmi[180] in grado di ricostruire i fatti sulla base degli indizi che in precedenti casi hanno svolto un ruolo nella ricerca[181].

Il programma ALIBI, di fronte a uno specifico crimine, è in grado di prevedere le giustificazioni del comportamento che fornirebbe l'accusato. Impersonando il sospettato, esso produce delle osservazioni e delle spiegazioni alternative rispetto

---

[177] F. Donati, *Intelligenza artificiale e giustizia* in *Rivista Associazione italiana dei Costituzionalisti*,N°: 1/2020,

[178] In argomento cfr. S. Governatori, M. Maltagliati, G. Marliani, G. Pacini, V. Pilla, *Come calcolare gli assegni di mantenimento nei casi di separazione e divorzio*, Milano, 2009, spec. 145 ss.

[179] C. Castelli, D. Piana, *ult.op.cit*. p.70

[180] Nello specifico, STEVIE e ECHO. In argomento cfr. Nissan, *Digital technologies and artificial intelligence's present and foreseeable impact on lawyering, judging, policing and law enforcemen*t, Springer-Verlag, London, 2015.

[181] J. Nieva-Fenoll, *Intelligenza artificiale e processo*. Giappichelli, Torino, 2019, pp. 14 et seq.

alle accuse mosse nei suoi confronti[182]. Il *software* scompone le azioni coinvolte nella sequenza criminosa separandole dalle loro connotazioni morali o legali, ad esempio "rubare" può essere interpretato come un semplice "prendere" in alcune circostanze, dunque non necessariamente oggetto di rimprovero. ALIBI simula una spiegazione in base alla quale l'atto del "prendere" risulti legittimo rispetto alla condotta dell'accusato[183].

Grazie al *Data Mining*[184] è stato inoltre possibile individuare i luoghi in cui è più probabile rinvenire tracce di delitto: specifici programmi, sulla base di dati provenienti da diverse scene del crimine, elaborano ipotetici luoghi del delitto, aiutando la polizia a indirizzare la ricerca nei luoghi dove le prove possano essere trovate con maggior probabilità[185]. Gli algoritmi sfruttano reti neurali "addestrate" su dati raccolti in precedenti scene del delitto, registrando indici di efficacia pari al 68%, superiori rispetto alle capacità di previsione dell'essere umano[186].

Anche nell'ambito dell'elaborazione delle argomentazioni sono stati progettati programmi in grado di fornire in un brevissimo lasso di tempo una serie di argomenti a favore e contro, supportandoli con riscontri documentali[187]. In secondo luogo, sono stati elaborati numerosi programmi di analisi di testi complessi, come atti processuali e allegati, capaci di collocare i fatti allegati nel corretto contesto giuridico e giurisprudenziale. I programmi sono stati inoltre in grado di rilevare il ragionamento seguito dalle giurie nelle decisioni, permettendo di individuare il modello mentale seguito dai giurati[188].

---

[182] E. Nissan, *Digital technologies and artificial intelligence's present and foreseeable impact on lawyering, judging, policing and law enforcement*, Springer-Verlag London 2015, cit., p. 11

[183] *Supra* nota 180

[184] Si intende un insieme di tecniche progettate per meglio intendere e interpretare i dati raccolti.

[185] J. Nieva-Fenoll, *ult. op. cit*., p.15. Vedere sul punto anche Adderley R, Bond JW, Townsley M. *Predicting Crime Scene Attendance* in *International Journal of Police Science & Management*. 2007, pp. 312-323.

[186] J. Nieva-Fenoll*, op. cit.*

[187] *Supra* nota 181, p.16

[188] Johnson- Laird P.N., *Mental models and probabilistic thinking,* in 50 *Cognition*, 1994, pp. 191 et seq.

Per supportare le argomentazioni sono stati ideati programmi[189] come ROSS INTELLIGENCE[190], capace di individuare gli argomenti più solidi da utilizzare nella controversia[191] e RAVEL LAW, strumento utilizzato negli USA a supporto dei legali, progettato per individuare le argomentazioni con maggiori *chances* di essere accolte in giudizio. L'algoritmo ricerca le decisioni e le citazioni di ogni singolo giudice, l'argomento più persuasivo per il giudice, le sue opinioni e i suoi scritti.

HYPO, sviluppato presso L'Università del Massachusetts, analizza le situazioni problematiche relative alle controversie sui segreti commerciali, recupera i casi legali rilevanti dal suo database e li trasforma in ragionevoli argomenti legali[192].

Programmi come ARGUMED e CATO[193] facilitano il lavoro di ricerca e strutturazione delle argomentazioni giuridiche a supporto di conclusioni, presentando modelli argomentativi estrapolati da precedenti casi giudiziari.

Risulta evidente che l'intelligenza artificiale può facilitare il lavoro di "persuasione" permettendo di raccogliere in modo molto più celere le informazioni necessarie. Non bisogna trascurare tuttavia il fatto che l'attività argomentativa richiede un lavoro di persuasione che non sempre dipende da variabili prevedibili[194].

4.1.2 L'Estonia e il progetto di un giudice robot

L'Estonia, da sempre Paese all'avanguardia in materia di *e-governance* e innovazione tecnologica, ha ideato nell'ambito di un ampio progetto di sviluppo

[189] Per altri programmi simili vedere Nissan, *Digital technologies and artificial intelligence's present and foreseeable impact on lawyering, judging, policing and law enforcement*, Springer-Verlag London 2015, cit., pp. 8- 9

[190] Per individuare il precedente o il principio di diritto che supporta al meglio un argomento, l'algoritmo effettua una ricerca per periodo e giurisdizione oggetto di interesse, fornendo i passaggi e i casi che più si avvicinano alle richieste del soggetto.

[192] Nissan*, ult. op. cit.*, p. 10

[193] Aleven, V. and K. Ashley. *Evaluating a learning environment for case-based argumentation skills*. *ICAIL ,*1997.

[194] J. Nieva-Fenoll*,ult. op. cit.*, p. 17

digitale un “giudice robot” in grado di svolgere funzioni giudiziarie per le controversie di valore inferiore a settemila euro. Il fine perseguito è alleggerire la mole di lavoro dei giudici per consentire a questi ultimi di concentrarsi sui casi più complessi.

Per rendere possibile la realizzazione di questo progetto, ancora in fase di sviluppo, sono stati utilizzati numerosi database legali con i quali l’algoritmo è stato “allenato” a produrre opinioni su determinate dispute[195].

Sulla base dei documenti processuali caricati sulla piattaforma dai ricorrenti, l’algoritmo è in grado di analizzare la normativa pertinente e gli atti rilevanti, emettendo una decisione sulla questione. È tuttavia prevista la possibilità di appellare la decisione richiedendo l’intervento di un giudice umano.

4.2 Profili critici e opportunità

Se impiegati in funzione di mero ausilio all’attività umana per le operazioni ripetitive e che non necessitano di valutazioni critiche, gli algoritmi possono fornire un contributo unico per lo sviluppo e l’efficienza del processo: gli avvocati possono utilizzare tale tecnologia per fornire ai loro clienti consigli più informati grazie a una valutazione empirica e sistematica delle probabilità di successo di una procedura, nonché incoraggiare la conclusione di transazioni che, qualora necessario, consentano di evitare un processo lungo e costoso[196].

Possono essere utilizzati per ridurre sensibilmente i tempi e i costi della giustizia, affiancare l’attività di cancellieri e giudici per le attività meccaniche (come la ricerca dei precedenti) e per la predisposizione di tabelle volte alla determinazione di importi monetari nell’ambito di giudizi civili (assegni alimentari,

[195] R. Banerjee, *Estonia develops "robot judge"*, New statesman, Vol.148(5474), , 2019, p.S5
[196] CEPEJ, *Carta etica*, par. 96

indennità compensative, risarcimento delle lesioni personali, indennità di licenziamento, ecc.)[197].

Tuttavia, il tema dell'applicazione degli strumenti di AI in ambito processuale ha sollevato problemi etici, di responsabilità e di rispetto dei diritti fondamentali dei soggetti coinvolti. Come osservato da Cathy O'Neil[198]: «Gli algoritmi ripetono le nostre prassi del passato, i nostri modelli, automatizzano lo status quo. Sarebbero straordinari se vivessimo in un mondo perfetto, ma non lo abbiamo[199].»

Siamo ancora lontani dalla possibilità di elaborare macchine capaci di riprodurre le infinite e complesse funzioni umane[200], *in primis* l'intuito, tanto più in fase di giudizio, settore che necessita più di ogni altro del coinvolgimento di fattori umani e psicologici per il lavoro di interpretazione (supportato da ipotesi anche adottate dalle diverse discipline delle scienze sociali)[201]. Gli algoritmi non esauriscono la peculiarità della condizione umana, che si sostanzia nella sua capacità di eccedere la massa dei dati, che sono mezzi e non scopi[202]: non è possibile delegare agli algoritmi dei procedimenti mentali che solo l'attività cerebrale dell'uomo è in grado di elaborare.

È necessario dunque stabilire dei confini oltre i quali non lasciar penetrare tali strumenti, preservando e valorizzando il ruolo certamente insostituibile dell'essere umano: si corre altrimenti il rischio che l'efficienza legale prevalga sull'equità[203].

---

[197] *Ibidem*

[198] Matematica, scienziata di dati e autrice americana.

[199] C. O' Neil, *L'era della fede cieca nelle grandi masse di dati deve finire*, Ted Talks, 2017 in https://www.ted.com/talks/cathy_o_neil_the_era_of_blind_faith_in_big_data_must_end/transcript?language=it

[200] "Dell'immaginazione; della capacità di dare vita a processi creativi; della coscienza, intesa secondo la teoria dell'informazione integrata; di creatività, emozioni e ispirazione, frutto dell'azione degli ormoni. E anche il beneficio del dub- bio, con il correlato senso di curiosità, e la sana consapevolezza di sapere di non sapere sono caratteristiche che contraddistinguono l'umano e la sua ricerca di senso, le quali mal si attagliano ai ragionamenti dell'AI". V. Manes, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in *Discrimen*, 2020

[201] L. De Renzis, *ult. op. cit*., p. 142

[202] B. Romano, *ult. op. cit*., p. 62

[203] *Ibidem*

4.2.1 La possibilità di utilizzare le determinazioni dell'algoritmo nel processo penale

Il fascino della perfezione degli automatismi non ha risparmiato l'ambito penale: il potere degli algoritmi sembra assurgere ad antidoto contro l'incertezza e l'imprevedibilità che inevitabilmente caratterizzano le dinamiche umane. Semplicemente inserendo una serie di dati in un programma, è ormai possibile calcolare il rischio di recidiva del condannato in un certo lasso temporale, prevedere dove, quando e chi commetterà un crimine e valutare l'attendibilità di una prova scientifica, per citare alcune possibilità.

In Italia, l'articolo 220, comma 2, del Codice di procedura penale vieta espressamente il ricorso a perizie per stabilire l'abitualità o la professionalità nel reato, la tendenza a delinquere, il carattere e la personalità dell'imputato e, in generale, le sue qualità psichiche indipendenti da cause patologiche.

Di tutt'altro avviso si è rivelata la giurisprudenza statunitense, che ha dimostrato la possibilità di integrare progressivamente gli algoritmi predittivi in campo penale (sia in fase di *pre-trial* che in fase di modulazione della sentenza[204]) con i rischi che questo comporta, i*n primis* in termini di lesione del diritto di difesa del soggetto.

In materia penale, proprio in virtù delle peculiarità e delle garanzie che devono essere assicurate per evitare un nocumento dei diritti, soprattutto in termini di equo processo e delle libertà fondamentali, è necessario svolgere un'accurata analisi dei costi e dei benefici che comporta l'introduzione di tali strumenti.

Come meglio si approfondirà nel corso della trattazione[205], sono molteplici le questioni sollevate: *in primis* il problema dell'opacità dell'algoritmo, che non consente di comprendere pienamente il meccanismo di funzionamento dello stesso e l'*iter* che ha condotto al risultato determinato; il funzionamento dell'algoritmo è coperto dal segreto industriale alla luce della normativa sulla tutela della proprietà intellettuale del s*oftware*, e ciò non permette alla controparte di vagliare gli elementi

[204] *State v. Loomis*, 881 NW 2d 749 (Wis 2016); vedi *infra* par. 7
[205] *Infra* Capitolo IV

su cui si basa il risultato, determinando un *vulnus* anche dell'obbligo di motivazione della decisione del giudice.

In secondo luogo, a destare preoccupazione è lo stesso meccanismo di funzionamento degli algoritmi: il *machine learning* contiene *in nuce* un problema di indeterminatezza e imprevedibilità, poiché nemmeno i programmatori più esperti spesso sono in grado di spiegare perché l'algoritmo produce quello specifico risultato.

Ulteriore profilo di rischio è dato dalla violazione del principio della responsabilità penale personale e del diritto a una sentenza individualizzata[206]. Altra critica mossa agli strumenti di automazione è la loro apparente neutralità: come si analizzerà nel corso del Capitolo V, numerosi studi[207] hanno dimostrato che i risultati dell'algoritmo sono viziati da *bias* discriminatori. A tal proposito, sono stati definiti delle "opinioni codificate" [208] proprio perché, volontariamente o involontariamente, riproducono i pregiudizi dei loro programmatori.

Appare evidente che il coinvolgimento di tali nuove tecnologie nel giudizio penale richiede maggiori cautele che in altri ambiti, nonché una tutela rafforzata dei principi contenuti nell'art. 6 CEDU in materia di equo processo, garantendo sempre il diritto di accesso all'algoritmo e il diritto di essere informati della logica alla base delle decisioni sottese agli algoritmi[209].

4.3 Processo cognitivo del giudice e limiti intrinseci dell'algoritmo

La *technology disruption* sta contribuendo a ridisegnare il ruolo del giudice nel processo: nell'attuale scenario tecnologico «si registra una tendenza crescente alla

[206] *Infra* cap. V, para. 5
[207] *Infra* par. 7
[208] C. O' Neil, *Armi di distruzione matematica. Come i Big Data aumentano la disuguaglianza e minacciano la democrazia*, Bompiani, Firenze, 2017
[209] CEPEJ, *Carta Etica*

digitalizzazione dell'amministrazione della giustizia e alla sostituzione del lavoro dell*'homo juridicus* con il *software*[210]».

Attualmente l'algoritmo viene utilizzato come strumento di supporto alle determinazioni del giudice in fase decisionale: il compito di tali strumenti è aiutare i giudici a prendere decisioni minimizzando il rischio di errori e di indebite influenze di fattori emozionali o esterni[211]. L'uso di tali strumenti si muove su un crinale sottile, quello tra mero supporto alle valutazioni e strumento delegato a decidere sulla libertà o reclusione del soggetto[212]. Soprattutto per quanto concerne la valutazione del rischio, i fattori da considerare nella modulazione della sentenza o della misura cautelare sono molteplici: non solo giuridici, ma anche sociali e psicologici.

In giudizio, applicare la norma non sempre assicura il *fare giustizia*[213]: in alcuni casi è richiesta un'attività d'interpretazione[214] in grado di andare al di là del significato letterale del testo, qualora conduca a un risultato ingiusto. È fondamentale saper operare un bilanciamento di valori che contemperi tutte le specificità del caso concreto: per esempio, un giudice potrebbe disporre il rilascio su cauzione di un'autrice di reato a rischio di recidiva sulla base di una gerarchia di valori, attribuendo maggiore importanza al suo ruolo di madre e di protettrice dei suoi figli, mentre l'algoritmo sarebbe in grado di determinare il rischio di recidiva

---

[210] L.D'Agostino, *Gli algoritmi predittivi per la commisurazione della pena* in *Diritto penale contemporaneo* ,2019, pp. 354-373

[211] Per quanto anche i giudici siano influenzati nelle loro decisioni da numerosi fattori, l'ingresso dell'IA nel giudizio non eliminerà il problema degli *implicit bias* poiché, come già evidenziato e come si approfondirà nel corso della trattazione, i programmatori possono replicare i loro pregiudizi nel *software,* contribuendo a una cristallizzazione degli stessi.

[212] fenomeno già osservato negli ordinamenti statunitensi, *infra* par.7

[213] C.V. Giabardo, *Il giudice e l'algoritmo (in difesa dell'umanità del giudicare)*, in *Giustizia Insieme* online*,* 2020 su https://www.giustiziainsieme.it/it/scienza-logica-diritto/1224-il-giudice-e-l-algoritmo-in-difesa-dell-umanita-del-giudicare

[214] Come afferma Bruno Romano «Il diritto ha un senso e una concretizzazione nelle relazioni tra gli esseri umani, esposte al sapere parziale e al dubbio, che configurano il pensiero, le intenzioni e le scelte nelle relazioni tra le persone del dialogo. Non ha invece alcun senso nominare il diritto nelle connessioni di dati, nella successione delle operazioni degli algoritmi, estranee, in ogni loro fase, alle intenzioni ed alle scelte.» B. Romano, *op. cit.,* p.31

con maggiore precisione, ma non sarebbe in grado di operare una simile gerarchia di priorità[215].

L'attività decisoria umana è complessa e non sintetizzabile in un linguaggio matematico: il giudice, tramite un ragionamento induttivo e deduttivo, effettua una analisi e una sintesi degli elementi del giudizio e perviene alla sentenza sulla scorta del proprio bagaglio di conoscenze, esperienze, studi e capacità di analisi. Giudicare è una combinazione «di conoscenza, formulazione e verifica di ipotesi e anche interferenza di emozioni umane per adattare la giustizia al caso concreto».[216]

Al fine di meglio comprendere il processo decisionale del giudice, si analizzeranno di seguito i due principali modelli di *judging* elaborati dalla scienza: la concezione formalista e quella realista. Secondo la prima, i giudici applicano la legge ai fatti del caso in maniera logica e meccanica: il giudice è come una «*giant syllogism machine*»[217].

Secondo l'altra concezione, invece, i giudici seguono un processo intuitivo per giungere a conclusioni che solo in un momento successivo razionalizzano con una motivazione ragionata[218].

Nessuno dei due modelli appare soddisfacente nel descrivere il processo cognitivo alla base del giudizio. È necessario integrare il portato di entrambi i modelli per avere un quadro completo dei meccanismi decisori: le intuizioni giocano un ruolo chiave nella prima parte del processo, ma poi intervengono considerazioni più soppesate e complesse, meno automatiche, come "correttivo". Il processo è dunque duplice: induttivo, cioè spontaneo, veloce e automatico, in un primo momento, deduttivo, cioè richiedente un maggiore sforzo mentale, nel secondo; i processi mentali in questa fase sono «*deliberate, rule-governed, effortful, and slow*»[219]. Di conseguenza il primo momento propone risposte intuitive ai problemi giuridici nel momento in cui sorgono, in un secondo momento vengono

[215] CEPEJ, *Carta Etica*
[216] J.Nieva-Fenoll*, op.cit.,* , p. 46
[217] Guthrie, Chris; Rachlinski, Jeffrey J.; and Wistrich, Andrew J., "Blinking on the Bench: How Judges Decide Cases" (2007). Cornell Law Faculty Publications. Paper 917, p. 2
[218] *Ibidem*
[219] *Supra* nota 217, p. 8

valutate le qualità delle considerazioni, che possono essere corrette, ripensate o confermate[220].

I giudici hanno mostrato una maggiore tendenza verso l'approccio intuitivo, anche in virtù degli stimoli che ricevono in fase di giudizio. Ad esempio, spesso è stato osservato che i giudici sono vulnerabili al *bias* del "sennò di poi" (*hindsight bias*[221]): i giudici valutando i fatti dopo che sono accaduti, corrono il rischio che sia sovrastimata la prevedibilità di determinati eventi[222].

4.4 L'utilizzo dell'intelligenza artificiale nell'attività di prevenzione dei reati

Nel 2009 William J.Bratton, allora Capo del Dipartimento di Polizia di Los Angeles, durante un congresso affermava che «Very soon we will be moving to a Predictive Policing model where, by studying real time crime patterns, we can anticipate where a crime is likely to occur[223]».

Quel che un tempo sembrava un progetto avveniristico oggi è divenuto realtà: i limiti alle facoltà intuitive e deduttive dell'essere umano possono essere brillantemente superati dagli strumenti di intelligenza artificiale, come hanno dimostrato i *software* di polizia predittiva, impiegati dalle forze dell'ordine nelle attività di contrasto alla criminalità.

Il potenziale dei *Big Data* in termini di analisi e controllo sociale era stato già intuito ma solo di recente si è destato l'interesse nei confronti di strumenti analitici capaci di elaborare grandi quantità di dati e, sulla base di essi, effettuare previsioni

---

[220] *Ibidem*

[221] Definito come «la tendenza, da parte di un soggetto a conoscenza della conseguenza di un evento, a ritenere erroneamente che questi avrebbe previsto tale conseguenza» In argomento Hawkins, S. A., & Hastie, R. (1990). *Hindsight: Biased judgments of past events after the outcomes are known* in*. Psychological Bulletin, 107*(3), 311–327

[222] *Supra* nota 2013, p.24

[223] *A National Interoperable Broadband Network for Public Safety: Recent Developments: Hearing Before the Subcomm. on Comnmc'ns, Tech., & the Internet of the H. Energy & Commerce Comm*., 111th Cong. 20 (2009) (statement of William J.Bratton, Chief, Los Angeles Police Department)

a supporto della prevenzione del crimine[224]. Basti pensare che ad oggi, solo negli Stati Uniti ben centocinquanta dipartimenti di polizia ne fanno regolarmente uso.

I sistemi di *machine learning* sono stati efficacemente traslati nell'apparato investigativo e adattati per prevenire e prevedere la commissione di reati, coerentemente con la riedizione moderna di una consolidata tendenza all'utilizzo di strumenti di valutazione del rischio basati su calcoli statistico-attuariali[225].

Come si analizzerà nel corso della trattazione, negli ultimi anni gli strumenti di *predictive policing* si sono moltiplicati negli ordinamenti europei e d'Oltreoceano[226] come strumento di supporto delle attività investigative, pur non mancando di sollevare numerosi interrogativi circa l'opportunità del loro impiego in un contesto così sensibile per le libertà civili.

Nell'era dell'«asimmetria della sorveglianza»[227], occorre mettere nella giusta prospettiva pregi e difetti di questi strumenti, tanto demonizzati e esaltati al contempo.

Nel successivo sottoparagrafo si analizzeranno il funzionamento e le implicazioni sottese all'utilizzo di strumenti predittivi in fase di *policing*, ormai ampiamente diffusi in tutta l'Europa.

4.4.1 I *software* di polizia predittiva

Gli strumenti di polizia predittiva, come suggerisce lo stesso nome, sono impiegati dalle autorità per prevenire la commissione di certi tipi di crimini, determinando il luogo e il momento in cui è più probabile che saranno commessi[228].

---

[224] Perry, Walter L., Brian McInnis, Carter C. Price, Susan C. Smith, and John S. Hollywood. *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. RAND Corporation, 2013, p. 2

[225] L.D'Agostino, *Gli algoritmi predittivi per la commisurazione della pena* in *Diritto penale contemporaneo*, 2019, pp. 354-373

[226] Sul punto *cfr*. Ales Zavrsnik, *Big Data*, *Crime and social control,* Taylor and Francis, 2017, pp. 194 et seq.

[227] Frank Pasquale, *Paradoxes of privacy in an era of asymmetrical social contro*l, in Ales Zavrsnik, *ult. op. cit.,* p. 13

[228] A. Zavrsnik, *ult. op. cit.,*

Alla base di tali strumenti vi è l'assunto che il crimine sia un fenomeno *prevedibile*[229], in quanto i criminali agiscono in base a schemi comportamentali che si ripetono: sarà più probabile che il soggetto riproporrà non solo le modalità, ma anche le condizioni di tempo e luogo più simili a quelle che hanno determinato il successo di un precedente delitto[230].

Lo sviluppo dei modelli predittivi si è mosso lungo due direttrici principali: una spaziale (*dove* e *quando* è più probabile che sarà commesso un reato) e una personale (*chi* ha più probabilità di esserne autore o vittima)[231]. Tra i più diffusi *software* del primo tipo vi è PredPol[232], impiegato in numerosi dipartimenti di polizia statunitensi.

Sulla scorta degli elementi di localizzazione storica dei reati forniti dai rapporti di polizia, l'algoritmo ricerca i *patterns*[233] criminali e produce una mappa degli *hot*

[229] Nello specifico «crime is clustered in particular areas that usually can be explained as a function of certain environmental factors that create vulnerabilities for victims at certain times. (…) Event-based theories like the routine activities theory suggest that crime is likely to occur "when motivated offenders converge, suitable targets exist, and capable guardians are lacking". Place-based theories focus instead on vulnerabilities in the location as the reason for the criminal activity. These vulnerabilities can include simple factors such as poor lighting, lack of police surveillance, attractive victims, or easy escape routes, among many other possibilities.»; in Andrew Guthrie Ferguson, *Predictive Policing and Reasonable Suspicion,* Emory Law Journal 62, no. 2 (2012), p. 273-274

[230] Ales Zavrsnik, *ult. op. cit*., pp. 194 et seq.

[231] Mark Andrejevic, *Data collection without limits* in Ales Zavrsnik, *ult. op. cit.,* pag. 112 et seq.

[232] Sviluppato nel 2013 da una *start up* californiana, PredPol elabora i dati storici sulla criminalità e calcola con successo i luoghi a più alto rischio di commissione di un crimine. Il vantaggio principale è che questo programma non prende in considerazione né la razza né l'appartenenza etnica per fare previsioni, ma solo tipo di reato, data/ora del reato e luogo. PredPol suddivide i reati in due classi: crimini violenti (come omicidi, aggressioni e stupri) e crimini "minori" come la vendita e lo spaccio di stupefacenti e il vagabondaggio. In argomento *cfr*. C. O' Neil, *ult.op.cit*., p. 125 et seq.

[233] I modelli predittivi impiegati sono stati due: la *Near Repeat Theory,* secondo la quale una volta che è stato commesso un crimine in un certo luogo, è statisticamente più probabile che in quel luogo e nelle vicinanze saranno commessi tipologie di reato simili. Un'altra matrice predittiva utilizzata è rappresentata dal *Risk Terrain Model* (RTM) per il calcolo della vulnerabilità del territoriale. Vedere sul punto A. Guthrie Ferguson, *Predictive Policing and Reasonable Suspicion,* Emory Law Journal 62, no. 2 (2012), p. 277-284

spots[234] della città, cioè delle zone a maggior rischio, che verranno sottoposti a un pattugliamento rafforzato[235].

I dati presi in considerazione dall'algoritmo combinano i fattori di rischio ambientali e sociali[236] più frequentemente associati alla commissione di reati.

Ad esempio, per il reato di furto, al fine di calcolare la *crime vulnerability* di un luogo sono stati considerati fattori di rischio la presenza di:

i. Complessi residenziali;
ii. Supermercati;
iii. Fermate di mezzi pubblici;
iv. Banche o istituti di credito;
v. Ristoranti o esercizi commerciali;
vi. Scuole;
vii. Parchi[237].

Altri fattori rilevanti oltre a quelli ambientali possono essere l'ora del giorno, il giorno della settimana[238], la prossimità a eventi d'intrattenimento, la stagione[239], le

[234] Sulla base, ad esempio, di dati come: la quantità e tipologia di reati commessi in quel determinato luogo, il numero degli arresti, il numero di chiamate al 911, il tessuto economico- sociale che compone il quartiere, la prossimità a luoghi dove sono stati compiuti delitti in tempo reale.
[235] CEPEJ,*Carta Etica*, par. 120
[236] Per il calcolo della vulnerabilità del territorio si utilizza un modello chiamato *Risk Terrain Model* (RTM). In argomento Andrew Guthrie Ferguson, *The Rise of Big Data Policing*, NYU Press, New York, 2017, pp. 62-83

[237] J.M.Caplan, L.W. Kennedy, J.D. Barnum, E.L Piza, *Crime in Context: Utilizing Risk Terrain Modeling and Conjunctive Analysis to Explore the Dynamics of Criminogenic Behavior Setting, in Journal of Contemporary Criminal Justice*, 33(2), 2017, pp. 133 et seq.
[238] Ad esempio, se è giorno di paga o un giorno festivo.
[239] Ad esempio, durante l'estate, con la chiusura delle scuole, c'è maggiore probabilità che si registrino reati minori e furti con scasso.

condizioni dell'ambiente (ad esempio, strade fornite o meno di illuminazione) e dati relativi alla situazione economica e demografica dell'area oggetto di studio [240].

L'impiego di tali tecnologie mira *in primis* ad ottimizzare l'allocazione delle risorse per avere migliori *chance* di anticipare la commissione di un crimine[241], riducendo quanto più possibile l'incertezza che inevitabilmente caratterizza l'attività investigativa.

In secondo luogo, l'attività di *predictive policing* risponde non solo a esigenze di prevenzione sul territorio, ma anche a esigenze di protezione sociale e ricerca probatoria [242] connessa con l'individuazione del responsabile dei delitti [243]. Il risultato è che nelle zone individuate come "ad alto rischio" l'attività di sorveglianza della polizia aumenterà, e con essa il numero di interventi effettuati e il tempo impiegato a sorvegliare la zona. Da uno studio condotto dal Center for Evidence-Based Crime Policy presso la George Mason University, è risultato che gli interventi della polizia sono più efficienti nella prevenzione dei crimini quando «*they are proactive, use specific (as opposed to general) strategies, focus on small places (or groups operating in small places), and develop tailor-made solutions that make use of a careful analysis of local problems and conditions*».

Il risultato è che nelle zone individuate come "ad alto rischio" l'attività di sorveglianza della polizia aumenterà, e con essa il numero di interventi effettuati e il tempo impiegato a sorvegliare la zona.

Pur contribuendo a ridurre il rischio di commissione di reati, come si dimostrerà in seguito, le previsioni algoritmiche hanno un impatto diretto sulle comunità soggette a pattugliamento, spesso in termini di discriminazione e eccessiva criminalizzazione di determinati gruppi sociali e/o zone residenziali.

---

[240] Perry, Walter L., Brian McInnis, Carter C. Price, Susan C. Smith, and John S. Hollywood. *Making Predictions About Potential Crimes* In *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*, RAND Corporation, 2013, pp. 17-56.

[241] P Jeffrey Brantingham, *The Logic of Data Bias and its Impact on Place-Based Predictive Policing* in Ohio State Journal of Criminal Law, 2018

[242] D. Polidoro, *Tecnologie informatiche e procedimento penale: la giustizia penale "messa alla prova" dall'intelligenza artificiale* in *Archivio Penale*, 2020, n. 3, p. 8

[243] Grazie alla c.d attività di *"crime linking*" utilizzate per identificare l'autore del reato.

Vi sono numerosi punti deboli e dubbi di legittimità associati all'utilizzo di tali strumenti. In primo luogo, le pattuglie inviate nelle zone da sorvegliare tenderanno a investigare o interpretare con maggior sospetto tutto ciò che vedono perché la loro percezione è "viziata" dalle previsioni dell'algoritmo[244]: di fatto, nelle zone segnalate, *«police may feel additional license to investigate more aggressively. (...) The knowledge of being in an area of higher violence may alter the daily practice of officers, leading them to resort to physicial force more often.*[245] »

In secondo luogo, l'algoritmo contribuisce a esacerbare il sovra-pattugliamento nelle comunità a maggioranza afroamericana o ispanica: questo non fa che rinforzare dinamiche discriminatorie, al punto che l'algoritmo arriverà a ritenere corretta l'associazione razza-criminalità, inviando così la polizia solo nei quartieri a maggioranza di persone di colore[246]. A titolo di esempio, se un dipartimento utilizza un algoritmo in cui sono stati introdotti *biased data*, le previsioni di rischio possono condurre a maggiori arresti nei confronti di soggetti appartenenti a una minoranza, e di conseguenza si amplifica la probabilità che quella minoranza sia considerata pericolosa[247]: «*if police activity is predicted by race, then subsequent policing (and hence the costs of policing) will be unevenly allocated by race. The result is greater black exposure to arrest and incarceration*.[248]»

[244] «Police are trained to detect criminal activity and they look at the world with suspicious eyes. Individuals in the predicted area, innocent or guilty, will be seen with the same suspicious eyes.» Andrew Guthrie Ferguson, *The Rise of Big Data Policing*, NYU Press, New York, 2017, p. 62-83

[245] *Ibidem*

[246] *«*Under this analysis, predictive policing algorithms will learn less about crime in predominantly white areas and will report that there is less of a risk of future crime in those areas, while learning more about predominantly Black neighborhoods and indicating that more police personnel should be sent to those areas».

[247] M. Hamilton, *Predictive Policing through risk assessment* in J. McDaniel, K. Pease, *Predictive Policing and Artificial Intelligence,* Taylor and Francis, Milton Park, 2021.

[248] A.Z. HUQ, Racial Equity in Algorithmic Criminal Justice, in Duke Law Journal, 2019, p. 1077

Allo stesso modo «*Neighborhoods characterized by poor relations with police might underreport crime, such that they receive fewer policing resources in the future*[249].»

Infatti, un problema più profondo è legato alla qualità dei dati utilizzati dall'algoritmo: tali dati storici relativi ai crimini sono spesso incompleti e parziali[250], senza contare che possono celare *implicit bias* legati alla razza.

Uno studio condotto in materia dal Professor Jeffrey Brantingham [251] ha evidenziato che se il sospettato è un soggetto di colore, potrebbe essere maggiore il grado di sospetto nei suoi confronti. Mentre se il soggetto di colore è una vittima, potrebbe essere invece minimizzata la portata dell'evento lesivo[252]. Secondo uno studio condotto su PredPol dall' Human Rights Data Analysis Group «*Black people would be targeted by the algorithm at twice the rate of white people* [253]». Un altro studio [254] condotto sulla popolazione di Oakland ha evidenziato che «Quando l'algoritmo utilizza i dati della polizia per generare previsioni di crimini di stupefacenti a Oakland, l'algoritmo raccomanda di indirizzare il doppio delle risorse di polizia alle aree nere rispetto alle aree bianche, nonostante i reati di stupefacenti

[249] *Ibidem*

[250] «Certain crimes like murder, burglary, and auto theft tend to be consistently reported to authorities, while other crimes like sexual assault, domestic violence, and fraud tend to be unreported.» *Supra* nota 232, p.1146. In argomento anche *cfr.* Ferguson, A.G. *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement,* NYU Press, New York, 2017, pp.62-83

[251] Professore di antropologia alla UCLA

[252] «For example, if the implicit bias involves racial stereotypes, then police interactions with a young man of color would tend to produce outcomes that are against his interests. If that young man is the victim of a crime, the implicit bias works to minimize the significance of victimization.9 If the young man is the suspect in a crime, the implicit bias works to maximize his liability. If we consider how the implicit bias operates for individuals of a non-targeted group, then the significance of victimization is maximized for the victim, and the liability is minimized for the suspect.» In J. Brantingham, *The Logic of Data Bias and Its Impact on Place- Based Predictive Policing,* Ohio State Journal of Criminal Law, vol. 15, no. 2 (2018), p. 476

[253] In argomento Renata M. O'Donnell, *'Challenging Racist Predictive Policing Algorithms under the Equal Protection Clause'* (2019) ,94 (3) New York University Law Review 544, p. 561.

[254] K. Lum, W. Isaac*, To Predict and Serve?* In *Significance*, Oct. 2016, at 16

fossero ragionevolmente distribuiti equamente sia nelle aree bianche che in quelle nere.[255]»

Anche il *tipo di reato* da prevedere può impattare sull'effetto discriminatorio delle tecnologie predittive: le sparatorie tendono a concentrarsi nei quartieri più poveri, che spesso soffrono di problemi di povertà e discriminazione legati alla razza, di conseguenza il *target* predittivo della violenza sarà maggiormente distorto che nel caso di furti[256]. Ciò contribuisce a perpetrare una criminalizzazione della povertà[257].

La loro raccolta è inoltre spesso correlata a pratiche di polizia razziste, specialmente in termini di numero di arresti in proporzione, proprio perché afroamericani e ispanici hanno maggiore probabilità di essere fermati dalla polizia per pregiudizio più che per effettiva necessità (al punto da poter parlare di *disparate policing*).[258]

Infine, questo ampio affidamento alle tecnologie predittive potrebbe minimizzare o addirittura sostituire progressivamente il giudizio umano (si parla *automation bias)*[259].

La soluzione è dunque non riporre una fiducia cieca verso questi dispositivi, di cui spesso il meccanismo di funzionamento non risulta trasparente, e richiedere una maggiore regolamentazione e trasparenza nei dati immessi nell'algoritmo per

---

[255] A.Z. HUQ, Racial Equity in Algorithmic Criminal Justice, in Duke Law Journal, 2019, p. 1077

[256] A. Guthrie Ferguson, *The Rise of Big Data Policing*, NYU Press, New York, 2017, p. 62-83

[257] C. O' Neil, *ult. op. cit.*

[258] «The disproportionate policing of communities of color may stem, in part, from biases of officers. Empirical evidence demonstrates that "police officers - either implicitly or explicitly - consider race and ethnicity in their determination of which persons to detain and search and which neighbourhoods to patrol. Disparate policing may also stem from white civilians' implicit biases, which cause them to conceive of people of color as "more dangerous" in some way, and cause them to call the police more frequently to address people of color than they would for similar behavior of white people.» In *Supra* nota 230, pp. 555-556

[259] *Predicting crime, LAPD style*, The Guardian, 25 giugno 2014 in https://www.theguardian.com/cities/2014/jun/25/predicting-crime-lapd-los-angeles-police-data-analysis-algorithm-minority-report

evitare *outcome* distorti da pregiudizi[260]. Non bisogna dimenticare che gli *hot spots* di *social crime* non sono altro che *hot spots* di *social needs:* sarebbe più funzionale che l'intervento proattivo fosse diretto anche a supporto delle fragilità dei luoghi soggetti a controllo, oltre che a anticipare la commissione dei reati[261].

4.5 Diffusione degli strumenti predittivi in Italia: l'algoritmo *KeyCrime* e *XLAW*

Anche l'Italia ha portato avanti progetti nell'ambito delle tecnologie predittive a contrasto della criminalità: allo stato attuale vengono utilizzati diversi software di polizia predittiva, tra cui KeyCrime e XLAW.

Il primo, originariamente ideato dalla questura di Milano, rifacendosi alla tecnologia di *crime linking*, anziché prevedere gli *hot spots* in cui saranno commessi i reati, si concentra sull'individuazione delle persone che con più probabilità potranno commetterli. Il *softwar*e analizza le condotte seriali di alcuni criminali (già identificati o da identificare) e la sua efficacia è data dalla quantità e qualità di dati raccolti dalle persone offese, inseriti nel programma insieme agli altri elementi oggettivi del fatto[262]. Il programma è in grado di immagazzinare e analizzare fino a 12.000 informazioni per ciascun atto criminoso e compiere una duplice analisi: induttiva, per individuare gli aspetti comuni a eventi diversi e una di tipo deduttivo, fornendo una previsione[263].

---

[260] «In the policing context, the unthinking use of algorithmic instruments will reinforce historical race-based patterns of policing. This may occur because algorithmic predictions will vary depending on the quality of the training data used to construct the predictive function.» A.Z. HUQ, Racial Equity in Algorithmic Criminal Justice, in Duke Law Journal, 2019, p. 1076

[261] Andrew Guthrie Ferguson, *The Rise of Big Data Policing*, NYU Press, New York, 2017, p. 161-176

[262] C. Parodi, V. Sellaroli, *Sistema penale e Intelligenza Artificiale: molte speranze e qualche equivoco* in *Diritto penale contemporaneo*, n. 6/2019, pp.56-59

[263] M. Venturi, *KeyCrime, La chiave del crimine* in Profiling. I profili dell'abuso, 10 febbraio 2015 consultabile su http://eprints.bice.rm.cnr.it/id/eprint/10312

I dati [264] vengono analizzati e gli eventi sono classificati in base alle caratteristiche personali dell'autore del fatto: ciò permette di mettere a confronto i crimini commessi in una determinata area per individuare eventuali *pattern* riconducibili a un soggetto. Il programma è, inoltre, in grado di comparare le abitudini operative degli autori dei reati (espressioni utilizzate, orari dei fatti, modalità di allontanamento del luogo dell'evento etc.). Il Questore di Milano, Luigi Savina, ha affermato che «Attraverso l'utilizzo di questo sistema, la Questura di Milano è riuscita a contenere e a ridurre in modo significativo i fenomeni di rapina perpetrati in ambito bancario e commerciale, producendo un effetto di deterrenza; a ciò ha contribuito sicuramente anche l'individuazione di numerosi autori responsabili di questo efferato crimine.»[265]

XLAW è un progetto ideato dall'Ispettore Superiore di Polizia di Napoli, Elia Lombardo, basato anch'esso sulla ciclicità e stanzialità di alcuni reati: l'algoritmo permette di decodificare la strategia dei disegni criminosi, prevedendone l'insorgenza sul territorio[266]. Il modello supera sia il *crime linking,* sia la mappatura degli *hot spots.* Difatti, basando l'attività del programma sulla selettività e sequenzialità dei controlli in virtù delle previsioni elaborate dal modello, è possibile prevenire i crimini più efficacemente. I controlli sul territorio sono disposti in maniera selettiva e puntuale, registrando in tal modo una diminuzione di scippi, rapine, furti e borseggi[267]. È stato possibile, inoltre, ottimizzare l'allocazione delle risorse operative del corpo di polizia, migliorando anche la percezione di sicurezza del cittadino[268].

Il programma considera numerosi fattori, tra cui il numero dei delitti avvenuti in un determinato luogo, il numero di esercizi commerciali, la presenza di istituti di

[264] Chiaramente solo quelli raccolti nel rispetto della disciplina codicistica potranno essere utilizzati.
[265] M. Venturi, *KeyCrime, La chiave del crimine* in Profiling. I profili dell'abuso, 10 febbraio 2015 consultabile su http://eprints.bice.rm.cnr.it/id/eprint/10312
[266] https://www.xlaw.it/presentazione/
[267] *Ibidem*
[268] G. Crupi, XLAW - *Innovazione strategica e tecnologica per la prevenzione dei reati predatori urban*i in Polizia Pubblica Sicurezza online, n. 69

credito, programmazione di eventi che determinano aggregazione sociale e condizioni climatiche, generando un *PCrime,* un indicatore della pressione esercitata dal crimine sul territorio in un dato momento[269].

4.6 Diffusione degli strumenti predittivi in alcuni Paesi europei: brevi cenni

Di seguito si presenterà brevemente la diffusione e l'attuale impiego degli strumenti predittivi in alcuni Paesi europei, con particolare riferimento all'esperienza di Inghilterra, Danimarca e Spagna.

Non bisogna commettere l'errore di ritenere che tali strumenti siano un fenomeno prettamente statunitense: in realtà sono diffusi nella maggior parte degli ordinamenti nazionali per i contesti applicativi più vari, anche se la loro diffusione non è certamente paragonabile a quella d'Oltreoceano [270].

Per dare contezza del fenomeno, basti pensare che nel 2016 la Commissione Europea ha finanziato un progetto per implementare i controlli alle frontiere che prevede l'elaborazione di un programma chiamato *iBorderCtrl*, avente il compito di controllare la veridicità delle dichiarazioni rilasciate da chi varca la frontiera europea. Una guardia "virtuale" dotata di tecnologia per il *lie detector* pone delle domande dirette all'interessato: se il *software* percepisce che il soggetto non sta dicendo la verità, dovrà subire un ulteriore controllo davanti a un agente di frontiera, altrimenti potrà passare il confine senza ulteriori controlli[271]. Attualmente viene sperimentato in alcuni paesi europei, tra cui Ungheria e Grecia. Lo scopo è potenziare i controlli alla frontiera tramite gli strumenti d'intelligenza artificiale.

[269] *Ibidem*
[270] Spielkamp, Matthias (2019) *Automating Society: Taking Stock of Automated Decision-Making in the EU,* BertelsmannStiftung Studies 2019, p.56
[271] *Ibidem*, p.37-38

4.6.1 Regno Unito e l'algoritmo HART

La sperimentazione più vicina all'esperienza statunitense[272] è quella del Regno Unito, dove la polizia di Durham, in collaborazione con l'Università di Cambridge, ha ideato un programma denominato *Harm Assessment Risk Tool* (HART).

L'obiettivo dell'algoritmo è predire la probabilità che il condannato compia reati nei successivi due anni, classificandolo in base a un rischio basso, moderato o alto[273]. Il risultato è utile per valutare se la persona può essere ammessa al programma di riabilitazione *Checkpoint,* alternativo all'esercizio dell'azione penale, possibile solo se il livello di rischio è basso o moderato[274].

Quanto al meccanismo di funzionamento sostanziale, la stima avviene sulla base di 34 parametri, 29 dei quali basati sui precedenti del soggetto e i restanti comprendenti il sesso, i precedenti penali, l'età e due tipi di codice di avviamento postale[275]. Questo ultimo parametro rischia di acuire la discriminazione dei soggetti che provengono dalle zone più degradate. Inoltre, possono innescarsi dei *feedback loops*, in quanto «Se la polizia risponde alle previsioni concentrando i propri sforzi sulle aree con codici postali a più alto rischio, più persone provenienti da queste aree verranno all'attenzione della polizia e saranno arrestate rispetto a coloro che vivono in quartieri a basso rischio e non mirati. Questi arresti diventano quindi

[272] Che utilizza l'algoritmo COMPAS, *infra* par. 7

[273] È considerato a rischio alto se la probabilità che commetta un crimine grave come un omicidio; il rischio è moderato se vi è probabilità che commetta un crimine minore. Il rischio è basso se non vi sono probabilità che commetta altri crimini. In argomento 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, n. 27, 2018

[274] M. Burges, *UK police are using AI to inform custodial decisions – but it could be discriminating against the poor* in https://www.wired.co.uk/article/police-ai-uk-durham-hart-checkpoint-algorithm-edit

[275] con il rischio di discriminare coloro che provengono da contesti socio-economici svantaggiati. In argomento 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, n. 27, 2018, p. 223.

risultati che vengono utilizzati per generare successive iterazioni dello stesso modello, portando a un ciclo di attenzione della polizia sempre maggiore.»[276]

La precisione nella stima del rischio di recidiva ammonta a circa il 62%, percentuale che scende a 52.7% in caso di alto livello di rischio di recidiva.

Come gli altri strumenti predittivi, anche l'HART ha sollevato dubbi in termini di pregiudizi razziali, opacità dell'algoritmo[277] con relativi problemi di trasparenza del processo decisionale, problemi di "profezie auto-avveranti" e risultati ingiusti che determinano l'inasprimento di dinamiche discriminatorie, senza contare il rischio di falsi positivi e falsi negativi [278] essendo risultati basati su analisi statistiche.

4.6.2 Danimarca e il progetto di profilazione dei minori a rischio per la *early detection*

Un caso che ha destato particolare clamore e suscitato forti critiche da parte dell'opinione pubblica è quello della Danimarca: nel 2018 l'amministrazione di Gladsaxe[279] aveva avviato la sperimentazione di un algoritmo per la profilazione dei minori provenienti da contesti socio-culturali difficili, per individuare i soggetti a rischio di abusi.

Il "modello Gladsaxe" [280] calcolava il punteggio di rischio in base a una serie di parametri relativi al contesto familiare di provenienza, tra cui: infermità mentale (3000 punti), disoccupazione (500 punti), appuntamenti dal medico mancati (1000

[276] *Ibidem*, p. 228
[277] Non è infatti possibile comprendere il meccanismo di funzionamento dell'algroritmo, e dunque anche le relazioni tra dati inseriti e risultato finale ottenuto. *Infra* capitolo IV, par. 3.1
[278] E' il caso di un soggetto individuato come a basso rischio, che poi commette un crimine considerato grave.
[279] Nel progetto erano coinvolte anche altre due città: Guldborgsund e Ikast-Brande.
[280] In un primo momento la Danimarca aveva suggerito di estendere il modello a tutto il Paese nell'ambito di un più ampio progetto di "*ghetto- plan*" avente l'intento di individuare i quartieri qualificabili come *ghetto* in cui sarebbero state introdotte speciali misure, tra cui l'applicazione degli *automated risk assessment system* per le famiglie con bambini. Il progetto è stato poi abbandonato.

punti) e divorzio[281]. Le valutazioni erano effettuate e registrate, tra l'altro, senza che le famiglie ne avessero alcuna conoscenza, in violazione delle leggi sulla privacy.

Il progetto, dal sapore "Orwelliano", ha suscitato fortissime critiche e nel 2019 si è improvvisamente arrestato. In ogni caso, sono in corso altri esperimenti che prevedono l'impiego di *Risk Assessment tool* nel settore *welfare:* nel 2020 l'Università di Århus ha annunciato che sta sviluppando un algoritmo di supporto decisionale per individuare i minori più vulnerabili, suscitando di nuovo le perplessità dell'opinione pubblica[282].

4.6.3 Spagna e l'algoritmo SAVRY

L'algoritmo SAVRY (Structured Assessment of Violence in Youth) è uno strumento di *Risk Assessment* impiegato dai servizi sociali per stimare il rischio di recidiva di comportamenti violenti nei ragazzi dai 12 ai 18 anni. Il risultato è funzionale all'inserimento in un piano di intervento e sostegno [283].

L'assunto alla base dell'algoritmo è la possibilità di prevedere l'insorgenza di comportamenti violenti in età adolescenziale, in quanto gli stessi sono associati a indicatori ben determinati, come disturbi della personalità, comportamenti aggressivi e contesto familiare di provenienza. Il responso non si basa sul calcolo di un punteggio di rischio, ma su una considerazione clinica di tutti i fattori di rischio in un determinato caso[284]. Gli elementi considerati sono 24 e sono tutti empiricamente associati al rischio di condotte violente.

---

[281] Spielkamp, Matthias (2019) *Automating Society: Taking Stock of Automated Decision-Making in the EU,* BertelsmannStiftung Studies 2019, pp.50-51 in www.algorithmwatch.org/automating-society

[282] *Ibidem*

[283] *Supra* nota 263, p.122; in argomento *cfr*. Joanna R Meyers and Fred Schmidt, 'Predictive Validity of the Structured Assessment for Violence Risk in Youth (Savry) with Juvenile Offenders' (2008) 35 Crim Just & Behavior 344

[284] Monica Gammelgård , Anna-Maija Koivisto , Markku Eronen & Riittakerttu Kaltiala-Heino (2008) *The predictive validity of the Structured Assessment of Violence Risk in Youth (SAVRY) among institutionalised adolescents,* The Journal of Forensic Psychiatry & Psychology, 19:3 p. 353

Tali elementi sono divisi in 3 gruppi:

1. la storia e i precedenti del soggetto;
2. il contesto sociale di provenienza;
3. i fattori individuali e clinici.

A titolo di esempio, sono considerati fattori di rischio[285]:

- I precedenti del soggetto
- Casi di violenza domestica
- Precedenti dei genitori o del tutore
- Episodi di autolesionismo
- Profitto scolastico carente
- Impulsività
- Attitudine negativa
- Problemi di gestione della rabbia
- Deficit di attenzione
- Mancanza di supporto sociale

L'algoritmo mette in rapporto la valutazione finale con 6 "fattori di protezione"[286], determinando se il rischio è alto, moderato o basso a cui si affianca il parere di un esperto.

Studi condotti sulla validità dell'algoritmo hanno confermato l'affidabilità di questo strumento[287]. Come si cercherà di evidenziare nel prossimo paragrafo, gli strumenti di *machine learning* coinvolti nella valutazione del rischio hanno

[285] T. Grisso, G. Vincent, D.Seagrave, *Mental Health screening and assessment in juvenile justice ,* The Guilford Press, NY-London, 2005, pp. 311 et seq.

[286] Sono considerati fattori di protezione quei fattor che determinano una riduzione del rischio di comportamenti violenti, come l'attitudine positiva, un forte supporto sociale, legami forti, capacità di resilienza, rendimento scolastico positivo.

[287] In argomento Monica Gammelgård, Anna-Maija Koivisto , Markku Eronen & Riittakerttu Kaltiala-Heino (2008) *The predictive validity of the Structured Assessment of Violence Risk in Youth (SAVRY) among institutionalised adolescents,* The Journal of Forensic Psychiatry & Psychology, 19:3

mostrato risultati discriminatori nei confronti di ragazzi di sesso maschile, stranieri o soggetti appartenenti a determinate minoranze[288].

4.7 Il modello statunitense: il ruolo dell'algoritmo COMPAS nel processo penale

Nel presente paragrafo si analizzerà il ruolo e il funzionamento dei *risk assessment tools* nel processo penale statunitense, impiegati non solo in fase di *policing* ma anche in fase di *pre-trial* e di s*entencing* vero e proprio. Le corti statunitensi, in particolare dell'Arizona, del Colorado, del Delaware, del Kentucky, della Louisiana, dell'Oklahoma, della Virginia, di Washington e del Wisconsin[289] hanno mostrato, negli ultimi anni, una forte propensione all'utilizzo degli algoritmi di valutazione del rischio, al punto da potersi definire una «*Algorithmic Criminal Justice*[290]».

Il crescente impiego di questi strumenti è dovuto alle recenti conquiste della statistica, negli ultimi decenni impiegata per prevedere il comportamento umano, soprattuto in campo penale. Particolarmente utile si è rivelato il contributo in fase di giudizio prognostico del *periculum* per l'applicazione delle misure cautelari, essendo un evento incerto nell*'an* e nel *quando*[291]. Il ricorso a variabili statistiche sembra neutralizzare, o quantomeno ridurre, l'incertezza legata a tali valutazioni.

[288] *Ibidem*
[289] P. Benanti, *Algoritmi con pregiudizi: il caso serio delle corti di giustizia USA,* 2017 in https://www.paolobenanti.com/post/2017/10/03/algoritmi-con-pregiudizi-il-caso-serio-delle-corti-di-giustizia-usa
[290] «Algorithmic criminal justice, as I define the term, is the application of an automated protocol to a large volume of data to classify new subjects in terms of the probability of expected criminal activity and in relation to the application of state coercion..» A.Z. HUQ, Racial Equity in Algorithmic Criminal Justice, in Duke Law Journal, 2019, p. 1060

[291] J. Nieva-Fenoll, *op. cit.,* p.52

Esistono più di 60 tipi diversi di strumenti di *risk assessment* attualmente conosciuti[292], che prendono in considerazione fattori di rischio *statici* e *dinamici.*[293] Ulteriore distinzione da effettuare è quella tra strumenti sviluppati direttamente dai governi statali e quelli elaborati dalle imprese private. Nel primo tipo rientrano algoritmi come quello sviluppato dallo Stato del Virginia che ha ideato un proprio *tool* da applicare in fase di *sentencing*[294]. In quelli di matrice privata rientrano invece il Level of Service Inventory – Revised (LSI-R)[295] e il COMPAS[296] (Correctional offender management profiling for alternative sanctions), sviluppato dall'azienda privata Northpointe e che nel 2016, nel famoso e controverso caso *State vs Loomis*[297], ha portato alla condanna dell'imputato a 6 anni di carcere, non senza aver sollevato numerosi dubbi di legittimità costituzionale[298].

4.7.1 Diffusione degli algoritmi nelle corti statunitensi a livello federale e statale

Negli ultimi anni i *risk assessment tool* sono stati riconosciuti come lo strumento chiave della *criminal justice bail reform* che ha interessato gli Stati Uniti. Tali strumenti producono delle stime considerate più accurate di quelle che possono

[292] *Ibidem.* Per citarne alcuni LSI-R – Level of Service Inventory - Revised ,LSI/CMI - Level of Service/Case Management Inventory , ORAS - Ohio Risk Assessment System ,Static-99 (for sex offenders/ offenses only), STRONG - Static Risk and Offender Needs Guide ,Wisconsin State Risk Assessment Instrument.

[293] I primi sono fattori che non cambiano nel tempo (ad esempio il sesso), i secondi possono variare (ad esempio l'età).

[294] Lo strumento, sviluppato dalla Virginia Criminal Sentencing Commission è funzionale all'individuazione di criminali a basso rischio per assegnare loro un trattamento sanzionatorio adatto. Da Kehl, Danielle, Priscilla Guo, and Samuel Kessler, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing,* 2017, Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School, p.11

[295] Utilizzato quale ausilio per il *sentencing* in alcuni Stati, tra cui il Colorado, la California, l'Iowa, l'Oklahoma e quello di Washington. Sul punto vedere M.Gialuz,*Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assesssment tools tra Stati Uniti e Europa* in Diritto penale contemporaneo, 2019 su https://archiviodpc.dirittopenaleuomo.org/upload/6903-gialuz2019b.pdf

[296] Usato ad esempio in California, Michigan e nello Stato di New York

[297] Corte Suprema del Wisconsin, *State v. Loomis*, 881 NW 2d 749 (Wis 2016)

[298] In termini di rispetto della *due process clause* e della *equal protection clause.*

effettuare i giudici contribuendo non solo a limitare *pretrial detention* inutili e incarcerazioni di criminali non violenti, ma offrendo a prima vista un correttivo alle decisioni potenzialmente viziate da pregiudizi, consapevoli o meno, dei giudici. Tuttavia, come si argomenterà, tali strumenti non fanno altro che amplificare e cristallizzare i summenzionati pregiudizi, inasprendo le disparità sociali.

Nel 2004 già 28 Stati utilizzavano strumenti di *risk assessment.* Le corti Supreme degli Stati hanno approvato l'utilizzo di questi strumenti e in alcuni casi lo hanno addirittura incoraggiato, come in Kentucky, Ohio, Oklahoma, Pennsylvania e Washington[299]. Oggi sono utilizzati praticamente in tutto il Paese e almeno in una contea in ogni Stato per un totale di oltre 1000 contee che ne fanno uso.

In diversi Stati la legislazione nazionale richiede espressamente ai giudici di valutare il rischio di commissione di reato o di mancata presentazione in udienza ricorrendo agli strumenti summenzionati. Interessante il caso dell'Idaho, in cui è permesso fare ricorso agli strumenti di valutazione algoritmica del rischio con la specifica che «*shall not be used until shown to be free of bias against any class of individuals protected from discrimination by law Pretrial risk assessment algorithms must be validated before use*.»[300]

I *national tools* più diffusi[301] sono:

- il Public safety assessment (PSA), impiegato ad esempio in Arizona, Kentucky e New Jersey in fase di *pre-trial,* per stabilire se il soggetto compirà un crimine o non si presenterà in giudizio. Lo Chief Regional District Judge, Karen Thomas, del Kentucky ha commentato «For a judge, the Public Safety Assessment provides an unbiased method of ensuring that individuals before the court are

[299] B. Harcourt, *Against prediction: Profiling, policing and punishing in actuarial age*, University of Chicago Press, Chicago,2007
[300] Legislature of the State of Idaho: House Bill No. 118. *Cfr. Mapping pretrial injustice: A community-driven database* in https://pretrialrisk.com/national-landscape/state-laws-on-rats/
[301] *Common Pretrial Risk Assessment* in *Mapping pretrial injustice: A community-driven database* in https://pretrialrisk.com/the-basics/common-prai/

afforded all of their constitutional rights, while also ensuring the safety of the community[302].» È impiegato in 59 contee (attraverso 20 Stati differenti) e 5 Stati per un totale di 56.3 milioni di persone potenzialmente coinvolte.

- Il Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), utilizzato in 11 giurisdizioni[303] attraverso 4 Stati.

- L' Ohio Risk Assessment System, impiegato in 43 contee (attraverso 11 Stati) e in 5 Stati.

- Il Virginia Pretrial Risk Assessment Instrument (VPRAI), impiegato in 16 contee e 1 Stato. In Virginia è utilizzato anche un altro programma particolarmente innovativo promosso dalla Virginia Criminal Sentencing Commission, il Nonviolent Risk Assessment (NVRA), per individuare soggetti coinvolti in crimini di droga e *property offenders* a basso rischio per applicare misure alternative alla detenzione[304].

Ben 11 Stati hanno sviluppato *State tools* (ad esempio Colorado, Florida, Maryland, Michigan, Minnesota, Oregon, e Washington) mentre per quanto riguarda lo sviluppo di *local tools* si consideri che oltre 50 contee hanno creato *tools* specifici per il loro territorio: in Kansas, California, Texas, Florida, Michigan e Maryland tutte le giurisdizioni hanno sviluppato questo tipo di strumenti[305]. Gli unici Stati in cui attualmente non è richiesto l'utilizzo degli strumenti di *risk assessment* in giudizio sono l'Arkansas, il Massachusetts, il Mississipi e il Wyoming.

---

[302] *21 Cities, States Adopt Risk Assessment Tool to Help Judges Decide Which Defendants to Detain Prior to Trial*, 2015 in https://www.arnoldventures.org/newsroom/more-than-20-cities-and-states-adopt-risk-assessment-tool-to-help-judges-decide-which-defendants-to-detain-prior-to-trial/

[303] Per citarne alcune: New York, Wisconsin, California, Florida's Broward County.

[304] De Keijser, J.W., Roberts, J.V. and Ryberg, J., *op.cit.*

[305] *Common Pretrial Risk Assessment* in Mapping Pretrial Injustice, in https://pretrialrisk.com/the-basics/common-prai/

4.7.2 L' Algorithmic Accountability Act

L' *Algorithmic Accountability Act*[306] è un disegno di legge del 2019 presentato negli USA su iniziativa di alcuni senatori che ha come scopo la regolamentazione degli algoritmi utilizzati dalle grandi compagnie: il fine ultimo è garantire che le suddette entità conducano valutazioni dei sistemi che prendono decisioni automatizzate (come quelli che impiegano l'intelligenza artificiale e il *machine learning*)[307], identificando e correggendo eventuali discriminazioni o pregiudizi nascosti all'interno di essi.

L'atto normativo si applicherebbe alle cosiddette *covered entity* definite come "*any person, partnership, or corporation that is subject to the jurisdiction of the Federal Trade Commission (FTC) and that either generates over $50 million per year, possesses or controls information on at least one million people or devices, or mainly operates as a data broker that sells or trades with consumer information*."[308]

Oggetto della regolamentazione sono nello specifico i sistemi decisionali automatizzati ad alto rischio[309], che includono quelli che possono contribuire a rinforzare pregiudizi o discriminazioni o che sono preposti a valutare e prevedere i comportamenti dei consumatori, influenzando aspetti sensibili della loro vita.

Le valutazioni che sarebbero richieste alle compagnie devono includere il funzionamento nel dettaglio del sistema, una valutazione dei suoi costi e benefici, la determinazione dei rischi per la sicurezza delle informazioni personali e illustrare le misure adottate per minimizzare tali rischi[310].

Il disegno di legge, con grande lungimiranza e certamente memore del dibattito europeo circa l'impatto degli algoritmi su alcuni diritti umani, evidenzia i rischi di

[306] US Congress. 2019. *Algorithmic Accountability Act*
[307] *Ibidem*
[308] *Ibidem*
[309] Tali sono considerati se sollevano problemi di privacy o sicurezza, coinvolgono informazioni personali di un significativo numero di persone o monitorano sistematicamente un luogo fisico e accessibile la pubblico.
[310] *US Congress. 2019. Algorithmic Accountability Act*

parzialità e di pregiudizi che possono nascondersi dietro ai risultati forniti dai sistemi decisionali automatizzati. Di fatto, come si affronterà in maniera più approfondita nel Capitolo V, i dati che vengono forniti al sistema decisionale possono riflettere i pregiudizi della società che descrivono, contribuendo a rafforzare discriminazioni ai danni di determinati gruppi sociali[311].

4.7.3 I Risk assessment tool per la valutazione della pericolosità sociale in fase di applicazione della misura cautelare e di formulazione della sentenza

La cultura della valutazione del rischio ha fatto parte del sistema americano sin dal 1930, ma per decenni tali valutazioni venivano effettuate clinicamente, cioè tramite il parere di un esperto.

I giudici incontrano numerose difficoltà nel compiere valutazioni di pericolosità sociale, proprio a causa dei numerosi fattori sociali e criminogeni coinvolti. Per questo negli ultimi decenni sono stati sviluppati sofisticati strumenti in grado di combinare fattori di rischio statici e dinamici, atti ad orientare le valutazioni del giudice nelle varie fasi del giudizio.

A differenza dei tradizionali strumenti di *risk assessment,* gli strumenti algoritmici di valutazione del rischio non sono costruiti sulla base di una comprensione teorica del comportamento criminale: i modelli predittivi sono basati sull'estrapolazione di *patterns* da *big data* e sull'uso di modelli statistici[312].

In sostanza, questi strumenti sono impiegati per effettuare *prognosi di pericolosità sociale*[313] e in particolare per determinare il rischio di recidiva del

[311] *Infra* capitolo V

[312] Schuilenburg, M. and Peeters, R. (eds.) (2020). *The Algorithmic Society: Technology, Power, and Knowledge*, Routledge, 2020

[313] «Risk assessment tools, in essence, use data regarding groups of people, a range of factors and weightings and human-inputted rules to predict an individual's future behaviour. Such instruments provide statistical predictions, typically comprising risk factors as predictors of violence or reoffending so that an individual is evaluated against these risk factors and 'scored' – the higher the score, the higher the risk and are used to score the risks of 'flight, rearrest, parole violation ... based on data from other people with characteristics similar' (Eck- house et al., 2019, p. 3).» Carolyn McKay (2020) *Predicting*

soggetto arrestato o incriminato, cioè la probabilità e la tipologia di futura condotta criminale[314]. Viene assegnato un punteggio di rischio per lo più combinando informazioni personali del soggetto con dati statistici di gruppi sociali e etnici usati come dataset di riscontro[315]. Il risultato è utile al giudice per decidere se concedere la libertà vigilata, per personalizzare i programmi di sostegno ai detenuti, per stabilire chi ha diritto alla libertà su cauzione e (impropriamente) anche per determinare la severità della condanna[316], a seconda della fase di giudizio in cui vengono utilizzati.

Ciò ha apportato numerosi vantaggi in sede processuale: questi strumenti permettono in primo luogo di reperire informazioni che sarebbero state, altrimenti, inaccessibili e di effettuare previsioni con un livello di precisione maggiore rispetto alle capacità di analisi dell'essere umano[317].

Allo stato attuale l'utilizzo attiene alle fasi di:

1. ***Policing****:* come già argomentato[318], gli strumenti di polizia predittiva come PredPol sono utilizzati per effettuare previsioni su luoghi e autori del reato.

2. ***Pretrial****:* un secondo uso attiene alla fase del *pre-trial*, quella in cui il giudice decide se applicare la misura di detenzione preventiva o rilascio su cauzione prima del processo. A titolo di esempio, in New Jersey e in molti altri Stati si utilizza il *Public Safety Assessment* (PSA) sviluppato dalla Arnold Foundation per stabilire se l'imputato si presenterà in tribunale, e se nel periodo di tempo intercorrente, commetterà un reato[319]. Se il soggetto è considerato a basso

---

risk in criminal procedure: actuarial tools, algorithms, AI and judicial decision-making, Current Issues in Criminal Justice, 32:1, 22-39

[314] S. L. Chanenson, J. M. Hyatt, *The Use of Risk Assessment at Sentencing: Implications for Research and Policy,* 2016, Villanova University Charles Widger School of Law

[315] S. Quattrocolo*, Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale "predittiva"* in *Cassazione penale* n .4/2019

[316] H.Fry, *op.cit*., p.64

[317] *Supra* nota 290, p. 1065

[318] Cap. III, par. 5

[319] M. Hill, *What goes into the algorithm behind New Jersey's bail reform?, NJ Spotlight news,* 13 settembre 2017 in https://www.njspotlight.com/video/goes-algorithm-behind-new-jerseys-bail-reform/

rischio viene rilasciato, altrimenti si applica la misura della custodia, sulla base del fatto che rappresenta un rischio per la comunità.[320]

3. ***Sentencing****:* recentemente i *risk assessment tool* hanno assunto un ruolo significativo anche nella delicata fase del giudizio. I giudici fanno sempre più affidamento sull'*outcome* dell'algoritmo per orientare le proprie decisioni in tema di determinazione della pena da applicare (in alcune giurisdizioni[321] l'utilizzo del risultato fornito da tali strumenti è addirittura fortemente incoraggiato dalla legge).[322]

Nonostante le indicazioni e le cautele su come tali strumenti andrebbero utilizzati, i giudici spesso ne fanno un uso *improprio*, modulando la gravità della pena sulla base del risultato dell'algoritmo e quasi sostituendoli alle *sentence guidilines* dettate in materia, che hanno invece lo scopo di standardizzare le condanne, limitando la discrezionalità dei giudici nella scelta del *range* di sentenze da applicare in relazione alla gravità del crimine commesso[323].

Rispetto al *pretrial*, questa è la fase più delicata del processo: il giudice deve stabilire come applicare la misura punitiva e la durata della stessa[324]. Vi è uno

---

[320] M. Stevenson, *Assessing Risk Assessment in Action*, 103 MINN. L. REV. 303, 319 20 (2018) «L'utilità risiede nel fatto che «Risk assessment has become a foundation for the bail reform movement by offering a substitute to a long-standing dependence upon monetary bail.Releasing more defendants who do not pose a substantial risk can alleviate the harms that money bail systems disproportionately wreak on poor and minority defendants.»

[321] Come in quella del Hampshire, della Pennsylvenia, dell'Arkansas e del Vermont. Ad esempio, in Oklahoma è imposto l'utilizzo di «assessment and evaluation instrument designed to predict risk of recidivism to determine eligibility for any community punishment».

[322] See ARK. CODE ANN. § 16-93-615(a)(1)(B) (2016) «The determination . . . shall be made by reviewing information such as the result of the risk-needs assessment to inform the decision of whether to release a person on parole by quantifying that person's risk to reoffend, and if parole is granted, this information shall be used to set conditions for supervision»; *Supra* nota 277, p. 1075

[323] J. Eaglin, The Perils of 'Old' and 'New' in Sentencing Reform (August 7, 2020). NYU Annual Survey of American Law, Forthcoming,

[324] D. Kehl, P. Guo, S. Kessler. 2017. *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*. *Responsive Communities Initiative, Berkman Klein Center for Internet & Society*, Harvard Law School, p.10

stretto legame tra recidiva e riabilitazione sociale. Se il candidato è valutato ad alto rischio, sarà considerato meno idoneo alla riabilitazione e la pena sarà più severa[325]. Viceversa, se il soggetto è considerato a basso rischio, il giudice potrà optare per una misura alternativa alla detenzione[326]. In tale fase «authorities need to consider which goals of punishment they are trying to achieve and how algorithmic tools could help maximize for those goals, if possible.[327]»

Il *Guiding Principles Report* (2011) del National Working Group ha specificato che i *risk scores* forniti non dovrebbero essere impiegati per mitigare o aggravare la sentenza di condanna, ma piuttosto per ridurre la recidività, proprio perché forniscono informazioni utili per determinare le esigenze riabilitative del condannato[328], separando lo scopo punitivo della sentenza dallo scopo di *crime reduction.*

Alcune *sentencing guidelines*[329] raccomandano di fare riferimento ai *risk assessment* soprattutto per i soggetti considerati a basso rischio[330], al fine di incoraggiare l'applicazione di misure alternative alla detenzione e ridurre le sentenze di incarcerazione.

Nel caso *Malenchik v Indiana*[331] la Corte Suprema dell'Indiana ha statuito (con particolare riferimento a programmi come LSI-R) che i *tools* devono

[325] Come nel caso *State vs Loomis*, in cui il giudice ha condannato Eric Loomis a 6 anni di prigione perch «*The risk assessment tools that have been utilized suggest that you're extremely high risk to reoffend*.» *Infra* par. 8

[326] «Tale assunto tuttavia lascia adito a molti dubbi. Non vi è infatti alcuna fondata evidenza scientifica che confermi gli effetti positivi della lunga incarcerazione sulla probabilità di recidiva dell'individuo.» [326] L. D'Agostino, *Gli algoritmi predittivi per la commisurazione della pena* in Diritto penale contemporaneo 2/2019, p. 361

[327] *Ibidem*

[328] Identificare le caratteristiche criminogeniche specifiche dell'autore del reato è utile durante la sua supervisione o condanna per diminuire la probabilità di future attività criminali.

[329] *Cfr. United States Sentencing Commission, Guidelines Manual*

[330] In Virginia a titolo di esempio è utilizzato il Nonviolent Risk Assessment (NVRA) per escludere gli imputati a basso rischio da alcune misure, tra quella detentiva.

[331] *Malenchik v. Indiana*, 928 N.E.2d 564 (Ind. 2010). Nello specifico «Having been determined to be statistically valid, reliable, and effective in forecasting recidivism, the assessment tool scores may, and if possible should, be considered to supplement and enhance a judge's evaluation, weighing, and application of the other sentencing evidence in the formulation of an individualized sentencing program appropriate for each defendant.»

essere utilizzati unicamente per stabilire se possono essere applicate misure come la libertà vigilata, se sospendere in tutto o in parte una pena, le possibilità di inserimento in un programma di correzione, se assegnare un delinquente a strutture o programmi di trattamento alternativi, valutare l'inserimento in programmi di correzione o il rilascio da custodia istituzionale[332].

Visto il crescente impiego di tali strumenti in fase di *sentencing,* l'American Law Institute ha proposto una riforma del *Model Penal Code* nella parte attinente al *sentencing* al fine di avallare l'utilità delle previsioni statistiche ma, al contempo, auspicandone un utilizzo trasparente[333].

4.7.4 L' algoritmo COMPAS

Il Correctional Offender Management Profiles for Alternative Sanctions (COMPAS) è uno strumento di valutazione attuariale del rischio di recidiva sviluppato dalla società privata Northpointe, Inc.[334] ed impiegato sia in fase di

[332] *Ibidem*

[333] L. D'Agostino, *Gli algoritmi predittivi per la commisurazione della pena* in Diritto penale contemporaneo 2/2019, p. 361; Si consideri che l'Idaho è stato il primo Stato a divulgare un atto legislativo contenete la necessità di assicurare la trasparenza, l'*accountability* e il diritto a una spiegazione del risultato prodotto dai *risk assessment tool*s: «All pretrial risk assessment tools shall be transparent, and: all documents, data, records, and information used by the builder to build or validate the pretrial risk assessment tool and ongoing documents, data, records, and written policies outlining the usage and validation of the pretrial risk assessment tool shall be open to public inspection, auditing, and testing; (b) A party to a criminal case wherein a court has considered, or an expert witness has relied upon, a pretrial risk assessment tool shall be entitled to review all calculations and data used to calculate the defendant's own risk score; and (c) No builder or user of a pretrial risk assessment tool may assert trade secret or other intellectual property protections in order to quash discovery of the materials described in paragraph (a) of this subsection in a criminal or civil case.» La disposizione si trova in https://legislature.idaho.gov/statutesrules/idstat/Title19/T19CH19/SECT19-1910/

[334] Il fatto che sia sviluppato da una compagnia privata non è affatto irrilevante, in quanto il *software* è coperto dal segreto industriale: di conseguenza non è possibile accedere ai meccanismi e alle logiche sottese al funzionamento del programma. Nel caso *State vs Loomis* il ricorrente ha lamentato una violazione al suo diritto di difesa e alla *due process clause* proprio sotto questo profilo. *Infra* par. 7.4

pretrial, che di *sentencing.* L'algoritmo è in grado di predire il rischio di recidiva e di pericolosità sociale degli imputati e viene impiegato per decidere circa l'entità e le modalità di esecuzione delle sanzioni penali (soprattutto in ordine alla possibilità di accedere a misure alternative alla detenzione) [335]. Tale strumento è utilizzato nello Stato di New York, Wisconsin, California, Florida's Broward County e in altre giurisdizioni.

Esso è in grado di valutare:

- Il rischio di recidiva violenta entro due anni;
- Il rischio di recidiva generale entro due anni;
- Il rischio di mancata comparizione in udienza (*pretrial risk*)[336].

L'algoritmo assegna un punteggio di rischio combinando le informazioni personali dell'imputato, ottenute tramite un questionario di 137 domande, con dati statistici di gruppi sociali ed etnici usati come *dataset* di riscontro per calcolare il rischio in caso di mancata detenzione [337]. Il giudice, sulla base dell'*outcome* algoritmico, stabilisce se applicare la custodia cautelare in fase di *pretrial* o, se impiegato in fase di *sentencing*, stabilisce se possono essere applicate misure come la libertà vigilata o altri programmi alternativi alla detenzione. Di fatto, il prospetto dell'algoritmo, anche se non dovrebbe, finisce per influenzare le determinazioni del giudice per quanto concerne la gravità della misura da comminare [338].

Il *software* considera fattori quali [339]:

---

[335] P. Zuddas, *Intelligenza artificiale e discriminazioni*, in *Liber amicorum* per Pasquale Costanzo, 16 marzo 2020.

[336] Funzionale a decidere se il condannato possa essere rilasciato su cauzione prima del processo

[337] S. Quattrocolo*, Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale "predittiva"* in *Cassazione penale* n .4/2019. Di conseguenza «Sul piano predittivo, quindi, lo strumento prevede il rischio di ricaduta violenta, senza tuttavia offrire una spiegazione di tale rischio, ma in rapporto al dato statistico.»

[338] I risk assessment tools "are intended nor recommended to substitute for the judicial function of determining the length of sentence appropriate for each offender. But such evidence-based assessment instruments can be significant sources of valuable information for judicial consideration.» *Malenchik v. Indiana*, 928 N.E.2d 564 (Ind. 2010)

[339] Sample-COMPAS-Risk-Assessment-COMPAS SCORE", vedi Northpointe, Practioners Guide to COMPAS, 2012, pp. 23 ss.; J. Nieva- Fenoll, *op. cit*., p. 58

1. Precedenti criminali della sua famiglia e dei suoi amici;
2. Criminalità nella zona di residenza (crimini commessi, quanti amici o familiari sono stati vittime di crimini nel quartiere, facilità di reperire sostanze stupefacenti, presenza di *gangs* sul territorio);
3. Situazione familiare (cioè se vive in una casa di sua proprietà o con la sua famiglia, o con amici);
4. Eventuale separazione dei genitori;
5. Livello di studi (ottenimento del diploma, votazione finale, rendimento scolastico generale, quante volte si è stati sospesi o esclusi, eventuali conflitti con insegnati ecc.);
6. Situazione lavorativa e finanziaria;
7. Situazione affettiva ed emotiva (ad esempio se è triste, annoiato o ha problemi di concentrazione, situazione di isolamento sociale per citarne alcuni);
8. Capacità di compromesso;
9. Tempo libero e ricreazione;
10. Abuso di alcol o sostanze stupefacenti;
11. Personalità criminale[340];
12. Propensione ideologica al crimine[341];
13. Carattere aggressivo o socievole;
14. Appartenenza del detenuto a una banda organizzata;
15. Precedenti arresti della persona;
16. Infrazioni disciplinari durante la detenzione.

Molti di questi elementi non solo non dimostrano alcuna propensione al crimine, ma sono anche fortemente discriminatori.[342] Spesso fattori come il livello di educazione, il quartiere residenziale e le condizioni economiche sono connesse all'etnia e alla classe sociale, svantaggiando sistematicamente le minoranze[343] e rinforzando il circolo vizioso incarcerazione - condizione sociale svantaggiata. Il risultato è che questi strumenti aiutano ad esacerbare le ineguaglianze che

[340] L'imputato deve esprimere il consenso o il dissenso rispetto a domande del tipo «I always practice what I preach» o «to get ahead in your life you must always put yourself first»

[341] Viene chiesto se ci si trova d'accordo con affermazioni come «a hungry person has a right to steal»

[342] J. Nieva-Fenoll, *op.cit.,* p. 56

[343] I più afflitti da condizioni socio- economiche svantaggiate sono afroamericani e ispanici. *Cfr*. De Keijser, J.W., Roberts, J.V. and Ryberg, J. (eds.) (2019). *Predictive Sentencing: Normative and Empirical Perspectives,* Bloomsbury Publishing.

determinano spesso comportamenti delinquenziali, incluso l'emarginazione sociale e la disoccupazione[344].

Inoltre, anche se la razza non è un fattore espressamente considerato, numerosi studi hanno dimostrato che il *software* ha mostrato risultati discriminatori nei confronti degli afroamericani e degli ispanici. Infine, si consideri che il COMPAS non calcola il rischio di recidiva *individuale*, tarato sul singolo caso specifico, ma effettua una *previsione statistica* generale, comparando il caso di specie a altri simili[345].

Anche se la decisione del giudice non può basarsi *unicamente* sul risultato dell'algoritmo, «il problema principale è che le percentuali di rischio di recidiva condizionano o almeno influenzano le decisioni giudiziarie sulla colpevolezza[346]», divenendo a tutti gli effetti uno strumento di supporto al *sentencing*.

L'impiego del COMPAS è stato contestato sotto diversi profili di legittimità costituzionale: il caso più noto è *State v. Loomis* che sarà analizzato di seguito.

4.7.5 Il caso State v Loomis

Prendendo le mosse dall'oramai emblematico caso Loomis, si andranno ad analizzare la portata e le implicazioni della storica sentenza della Corte Suprema del Wisconsin.

Nonostante la Corte abbia affermato la legittimità dell'utilizzo degli algoritmi predittivi in fase di giudizio*,* il caso pone in luce le criticità sottese all'ingresso degli algoritmi in fase di commisurazione della pena e di come questo possa determinare, in maniera silenziosa, un *vuluns* nei diritti fondamentali dell'individuo in ambito processuale.

---

[344] *Ibidem*

[345] https://www.giurisprudenzapenale.com/2019/04/24/lamicus-curiae-un-algoritmo-chiacchierato-caso-loomis-alla-corte-suprema-del-wisconsin/

[346] J. Nieva-Fenoll , *Intelligenza artificiale e processo,* cit., 140 ss

A essere minate sono in particolare il diritto dell'imputato di verificare l'accuratezza delle informazioni che hanno determinato la sentenza e il diritto a una sentenza individualizzata, entrambi fondamentali corollari del *fair trial*[347].

Come si vedrà, entrambi tali profili sono messi in discussione dalla natura del COMPAS: sotto il primo profilo a causa del segreto commerciale, sotto il secondo perché lo score del COMPAS, calcolato sulla base di dati di gruppo, è necessariamente a «*calculation for a generalized group*[348].»

i. I fatti

Nel 2013 Eric Loomis, alla guida di un veicolo precedentemente coinvolto in una sparatoria, veniva fermato dalla polizia e successivamente imputato per 5 capi d'accusa (tra cui messa in pericolo della sicurezza, tentativo di fuga od elusione di un ufficiale, guida di un veicolo senza consenso del proprietario e possesso di arma da fuoco).

Loomis si dichiarava colpevole solo per i due minori (tentativo di fuga od elusione di un ufficiale e guida di un veicolo senza consenso del proprietario). In fase di giudizio veniva utilizzato l'algoritmo COMPAS per la valutazione del rischio, che gli attribuiva un punteggio di rischio elevato sia per recidiva violenta, che generale. Il giudice, anche sulla base della previsione dell'algoritmo, stabilì di non concedere la libertà vigilata, condannando l'imputato a 6 anni di reclusione e 5 di *extended supervision*. Nel giustificare la dura pena comminata e la negazione della libertà vigilata, il giudice argomentava come segue:

[347] I problemi di *accuracy* e *individualization* della sentenza determinati dall'uso degli algoritmi erano stati già affrontati in due casi precedenti. Il primo problema è stato affrontato in *Gardner v. Florida,* in cui la Corte ha sottolineato il diritto dell'imputato di verificare l'accuratezza delle informazioni che hanno determinato la sentenza: «[t]he defendant ha[d] a legitimate interest in the character of the procedure which leads to the imposition of sentence even if he may have no right to object to a particular result of the sentencing process.». In *State v. Gallion* si ribadisce il diritto a una sentenza individualizzata come pietra angolare dell'equo processo.

[348] [348] K. Freeman, *Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State v. Loomis*, 18 N.C. J.L. & Tech. 75 (2016)

> «You're identified, through the COMPAS assessment, as an individual who is at high risk to the community. In terms of weighing the various factors, I'm ruling out probation because of the seriousness of the crime and because your history, your history on supervision, and the risk assessment tools that have been utilized, suggest that you're extremely high risk to re-offend.[349]»

La sentenza è significativa in quanto nella motivazione della sentenza il giudice richiama il risultato dell'algoritmo COMPAS: è dunque formalmente riconosciuto che la gravità della pena comminata è stata decisa anche sulla base del risultato fornito dall'algoritmo.

Per stabilire la durata della pena, tuttavia, bisogna contemperare una molteplicità di fattori (tra cui il rischio rappresentato per la comunità, la possibilità di riabilitazione, la funzione rieducativa della pena) e non solo il rischio di recidivare[350].

La pena aveva tra l'altro destato un certo clamore proprio per la gravità in rapporto ai fatti marginali per cui Loomis si era dichiarato colpevole. Loomis, lamentando la violazione del diritto a un equo processo, proponeva istanza di revisione al Tribunale circondariale. Durante l'udienza il Dr. David Thompson, chiamato a testimoniare dalla difesa in qualità di esperto, si pronuncia circa la legittimità dell'uso del COMPAS in giudizio. Il testimone argomentava che questi strumenti non dovrebbero essere utilizzati per stabilire la pena da comminare, con il rischio di «*over estimating an individual's risk and mistakenly sentencing them or basing their sentence on factors that may not apply*[351]». Inoltre, viene sottolineato che i tribunali non conoscono la modalità di funzionamento dell'algoritmo, né in che modo effettua la comparazione tra la storia del singolo individuo e quella del gruppo di popolazione di riferimento.

---

[349] *State vs Loomis,* 881 N.W.2d at 755
[350] H.Fry, *op. cit.*, p.66
[351] *State v Loomis,* punto 26 e ss.

Nonostante tali osservazioni, il tribunale rigettava l'istanza sostenendo che la sentenza sarebbe stata la medesima anche senza l'uso del COMPAS. Loomis impugnava davanti alla Corte d'Appello, che rimetteva la decisione alla Corte Suprema del Wisconsin.

ii. Profili di violazione della due process clause

Il ricorrente aveva lamentato la violazione della *due process clause* sotto diversi profili di illegittimità:

1. Violazione del diritto a essere valutato sulla base di informazioni accurate[352] per via della natura proprietaria dell'algoritmo;

2. Violazione del diritto a una sentenza individualizzata per via della previsione basata su dati statistici di gruppo generalizzanti e non attinenti all'individuo;

3. Lamentava infine l'incidenza del sesso tra i parametri di valutazione.

Sotto il primo profilo, Loomis lamentava l'inaccessibilità alle logiche sottese al funzionamento dell'algoritmo (coperto dal segreto commerciale in quanto proprietà intellettuale della Northepointe, Inc.), che rendeva impossibile comprende il meccanismo dell'algoritmo, come vengono selezionati i fattori di rischio e il peso attribuito a ciascun di loro, in danno al diritto di difesa del ricorrente e alla connessa possibilità di confutare il risultato. Il ricorrente non è stato in grado di comprendere con precisione in base a quale calcolo fosse stato prodotto il suo punteggio di rischio a causa della natura privata del COMPAS[353], fatto che ha comportato la violazione della *due process clause*[354].

[352] "[a] defendant has a constitutionally protected due process right to be sentenced upon accurate information." Travis, 347 Wis.2d 142, ¶ 17, 832 N.W.2d 491; Inoltre «The right to be sentenced based upon accurate information includes the right to review and verify information contained in the PSI upon which the circuit court bases its sentencing decision.» State v. Skaff, 152 Wis.2d 48, 53, 447 N.W.2d 84

[353] I. DE MIGUEL BERIAIN, *Does the use of risk assessments in sentences respect the right to due process? A critical analysis of the Wisconsin v. Loomis ruling* in Law, Probability and Risk (2018) 17, 45–53

[354] *Ibidem*

Per quanto concerne il secondo profilo di violazione, la difesa parte dal presupposto che il COMPAS parte da dati statistici di massa, esponendo l'imputato al rischio di subire gli effetti discriminatori della riconducibilità a un certo gruppo sociale anziché un altro[355].

Il COMPAS non stima la probabilità specifica che un individuo possa recidivare, produce invece una previsione basata su una comparazione delle informazioni relative all'individuo rispetto a simili dati di gruppo.

Presupposto di legittimità costituzionale è avere una sentenza determinata su quanto ha commesso l'imputato, non su quanto è stato commesso da altri che hanno delle similarità con lui. La valutazione del rischio deve essere effettuata ponderando le specificità del singolo individuo del caso concreto, non sulla base di previsioni statistiche e dati di massa appartenenti al passato ma con la pretesa di predire il futuro. Viene, inoltre, lamentato il rischio che un soggetto sia ritenuto più incline alla recidiva in virtù di fattori legati alla condizione sociale e di altri fattori che non dipendono direttamente dalle facoltà del soggetto.

Per quanto concerne l'ultimo profilo di doglianza, il genere sarebbe stato valutato come fattore criminogeno dall'algoritmo.

iii. La decisione della Corte suprema del Wisconsin

La decisione della Corte Suprema del Wisconsin segna un punto di svolta nel panorama giuridico statunitense per aver riconosciuto la legittimità degli algoritmi predittivi in ambito giudiziario, pur con la specifica che non possono costituire il *fattore determinante*[356] della pronuncia di condanna:

> «"[W]e conclude that if used properly, observing the limitations and cautions set forth herein, a circuit court's consideration of a

[355] S. Quattrocolo*, Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale "predittiva"* in *Cassazione penale* n .4/2019.

[356] «"[B]ecause the circuit court explained that its consideration of the COMPAS risk scores was supported by other independent factors, its use was not determinative in deciding whether Loomis could be supervised safely and effectively in the community.»

COMPAS risk assessment at sentencing does not violate a defendant's right to due process.»

I giudici, dunque, possono considerare il risultato dell'algoritmo, ma non possono fondare la decisione su di esso.[357] Nel caso di specie, la Corte ha rilevato che la valutazione del COMPAS ha avuto un ruolo marginale ai fini della decisione, essendo stati altri i fattori ad aver avuto un ruolo decisivo[358].

Quanto ai profili di doglianza della parte, la Corte ha disatteso tutte le censure del ricorrente: per quanto attiene alla trasparenza dell'algoritmo, il diniego di accesso a quest'ultimo non ha compresso il diritto di difesa dell'imputato in quanto «*Northpointe's 2015 Practitioner's Guide to COMPAS explains that the risk scores are based largely on static information (criminal history), with limited use of some dynamic variables (i.e. criminal associates, substance abuse)*[359]». Dunque, non viene riconosciuto il diritto di accesso all'algoritmo in capo all'imputato, poiché secondo i giudici sarebbe sufficiente la possibilità di supervisionare gli *input* immessi nell'algoritmo ed essere informato dell'*output* prodotto[360].

[357] Come si argomenterà, il confine tra le due attività è molto labile e spesso i giudici fanno grande affidamento su questi risultati, ritenendoli obiettivi e facendosi influenzare per la modulazione della sentenza.

[358] «Thus, the record reflects that the sentencing court considered the appropriate factors and was aware of the limitations associated with the use of the COMPAS risk assessment. Ultimately, although the circuit court mentioned the COMPAS risk assessment, it was not determinative in deciding whether Loomis should be incarcerated, the severity of the sentence or whether he could be supervised safely and effectively in the community. (…) As the circuit court explained at the post conviction hearing, it would have imposed the exact same sentence without it. Accordingly, we determine that the circuit court's consideration of COMPAS in this case did not violate Loomis's due process rights» Punti 109 e ss.

[359] Punto 54 della sentenza.

[360] C. Burchard, *L'intelligenza artificiale come fine del diritto penale? Sulla trasformazione algoritmi della società* in *Rivista italiana di diritto e procedura penale*, 2019, Vol. 62, N° 4,pp.1909-1942

La Corte parte dall'assunto che COMPAS utilizza solo dati pubblicamente disponibili (forniti dal casellario giudiziale) e dati forniti direttamente dall'imputato[361].

I giudici hanno ritenuto sufficiente per integrare il contraddittorio sulla prova, la possibilità di accedere al manuale d'uso del software fornito dalla Northpointe, Inc. e dei dati personali dell'imputato, che consentono la validazione dei risultati dell'algoritmo[362]. Sulla base delle informazioni contenute nel manuale, in cui si specifica che la valutazione predittiva dell'algoritmo è basata per la maggior parte su fattori di rischio *statici* attinenti alla storia criminale dell'individuo e solo in minor parte su fattori *dinamici* (ad esempio l'uso di sostanze stupefacenti) e dalle informazioni fornite dall'imputato stesso o estratte dal casellario, la difesa ha avuto modo di verificare l'accuratezza della valutazione[363].

In aggiunta, la Corte ha osservato che, anche se le determinazioni dell'algoritmo fossero inesatte, la competenza dei giudici compenserebbe tali inesattezze. Al contempo è stato ribadito che i giudici devono fare affidamento principalmente sulla loro discrezionalità e che il *presentence investigation report*[364] (PSI) deve osservare determinati obblighi di *disclosure* relativi allo strumento di valutazione del rischio, in particolare:

---

[361] «Loomis's risk assessment is based upon his answers to questions and publicly available data about his criminal history, Loomis had the opportunity to verify that the questions and answers listed on the COMPAS report were accurate.» L'argomentazione non ha convinto del tutto, in quanto informazioni accurate possono comunque produrre un risultato non attendibile. Non conta solo la qualità del dato inserito, ma anche la modalità di calcolo e il peso attribuito a ciascun fattore. In argomento Anne L. Washington, *"How to Argue with an Algorithm: Lessons from the COMPAS-ProPublica Debate,"* Colorado Technology Law Journal 17, no. 1 (2018): 131-160; Si determina inoltre un *vulnus* alla parità delle armi, in quanto non è possibile la validazione *ex post* del processo che ha portato a quel determinato risultato:minata la facoltà di contestare e criticare le prove contrarie..

[362] S. Quattrocolo, *ult. op. cit. p.1759*

[363] Punto 55

[364] La determinazione della pena nel processo è preceduta da una fase di istruttoria preliminare finalizzata a delineare il profilo socio- criminologico del reo. Prima di pronunciare la condanna il giudice attende il deposito del presentencing investigation report (PSI), in cui confluiscono gli elementi utili all'irrogazione della pena. Sul punto L. D'Agostino, *Gli algoritmi predittivi per la commisurazione della pena* in Diritto penale contemporaneo 2/2019, p. 360

> «Any PSI containing a COMPAS risk assessment must inform the sentencing court about the following cautions regarding a COMPAS risk assessment's accuracy: (1) th**e proprietary nature** of COMPAS has been invoked to prevent disclosure of information relating to how factors are weighed or how risk scores are to be determined; (2) risk assessment compares defendants to a *national sample,* but no cross-validation study for a Wisconsin population has yet been completed; (3) some studies of COMPAS risk assessment scores have raised questions about whether they **disproportionately classify minority offenders** as having a higher risk of recidivism; and (4) risk assessment tools must be **constantly monitored** and re-normed for accuracy due to changing populations and subpopulations. Providing information to sentencing courts on the limitations and cautions attendant with the use of COMPAS risk assessments will enable courts to better assess the accuracy of the assessment and the appropriate weight to be given to the risk score[365].»

Per evitare abusi in fase di commisurazione della pena, la Corte ribadisce, inoltre, che lo scopo di questi strumenti deve essere quello di individuare le esigenze specifiche del soggetto e il rischio di recidiva, ma non devono essere impiegati per determinare la gravità della pena[366].

Per quanto riguarda il diritto a una sentenza individualizzata, la Corte ha statuito che il diritto non è stato violato in quanto lo *score* del COMPAS è solo uno dei numerosi fattori considerati dal giudice nella formulazione della sentenza[367] e che si sarebbe profilata una simile violazione solo se lo *score* del COMPAS fosse stato un *fattore determinante* in fase di *sentencing*.

[365] *State v. Loomis*, 881 N.W.2d 749, 763–64
[366] «It is very important to remember that risk scores are not intended to determine the severity of the sentence or whether an offender is incarcerated.» Punto 17
[367] «COMPAS is merely one tool available to a court at the time of sentencing and a court is free to rely on portions of the assessment while rejecting other portions» Punto 92

Per l'ultimo profilo di illegittimità, secondo la Corte la difesa non aveva prodotto sufficienti prove a supporto della tesi; inoltre erano stati considerati numerosi altri fattori oltre al genere.

Nel complesso la sentenza non ha ricevuto valutazioni positive, tuttavia questi strumenti continuano a affiancare quotidianamente i processi decisionali dei giudici nelle più alte corti, in virtù dei pregi che si presume apportino al processo come baluardo contro i pregiudizi che possono condizionare il giudizio.

4.7.6 Ulteriori casi esemplificativi

L'utilizzo dei *risk assessment tool* in fase di *sentencing* è aumentato drammaticamente negli ultimi decenni. Il caso Loomis non è isolato: numerosi altri imputati sono stati condannati a una pena più severa a causa dell'*outcome* dell'algoritmo, la cui accuratezza e affidabilità è spesso sovrastimata, a maggior ragione per il fatto che nessuno conosce il meccanismo di funzionamento dell'algoritmo, protetto da segreto industriale. Nonostante gli ammonimenti, provenienti anche dalla Corte Suprema del Wisconsin nell'uso di tali strumenti, ha senso esemplificare in che modo i giudici ne fanno uso realmente.

A tal proposito può qui richiamarsi un ulteriore caso esemplificativo di quanto sopra rilevato. Il Signor Zilly, un uomo di 48 anni con precedenti per uso di metanfetamine, era stato imputato per il furto di un tosaerba e altri strumenti che intendeva rivendere. Nel 2013 si presenta davanti al giudice Babler, nella contea di Baron in Wisconsin, consapevole del fatto che i suoi difensori si erano già accordati con l'accusa per un patteggiamento.[368] Anche in tale caso il giudice utilizzò l'algoritmo COMPAS per calcolare il rischio di recidiva e per sua sfortuna l'algoritmo stimò un alto rischio. Il giudice decise così di ignorare l'accordo tra accusa e difesa e raddoppiare la pena da un anno di carcere a due in una prigione

[368] H. Fry, *op. cit.*, p.66 e ss.

statale, giustificando tale decisione con queste parole «*"When I look at the risk assessment, it is about as bad as it could be*"[369]».

La difesa, appellando la decisione, chiamò a testimoniare uno degli sviluppatori del COMPAS, che espresse preoccupazione per il fatto che le sentenze venivano comminate anche sulla base del risultato dell'algoritmo, che non era nato certamente per questo scopo ma con l'intento di ridurre la criminalità[370].

Anche sulla base di tale testimonianza, il giudice d'appello decise di ridurre la pena da 2 anni a 18 mesi[371] argomentando che «*Had I not had the COMPAS, I believe it would likely be that I would have given one year, six months*».

L'algoritmo, focalizzandosi solo sul passato, non aveva certo preso in considerazione tutti i cambiamenti e gli sforzi che Zilly stava facendo per reinserirsi nella comunità e per uscire dalla sua dipendenza dalle droghe.

Analoga situazione nel caso Christopher Drew Brooks, diciannovenne condannato per aver avuto un rapporto sessuale con una minore consenziente. Le *sentencing guidelines* suggerivano una condanna compresa tra i 7 e 16 mesi, ma fu utilizzato uno strumento attuariale di valutazione del rischio (diverso da COMPAS) che calcolò il coefficiente di rischio come elevato: la pena fu fissata a ben 18 mesi[372].

4.8 Implicazioni pratiche

Il caso di Eric Loomis è emblematico delle numerose criticità sottese all'uso degli algoritmi predittivi in fase di giudizio: il problema principale è l'affidamento che i giudici ripongono nei confronti di un programma di cui non è noto il

[369] J. Angwin, J. Larson, S. Mattu e L. Kirchner, *Machine Bias*, ProPublica in https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
[370] «I don't like the idea myself of COMPAS being the sole evidence that a decision would be based upon."»
[371] *Supra* nota 369
[372] *Ibidem*

funzionamento, minando in radice il diritto di difesa dell'imputato, impossibilitato a conoscere le basi sulle quali viene fornito il risultato.

La Corte del Wisconsin ha creato un precedente "pericoloso", aprendo le porte all'uso diffuso del COMPAS nei tribunali e affermando che il diritto di accesso all'algoritmo non è riconosciuto all'imputato e che questi non viene leso in alcun modo nel suo diritto a un *fair trial.*

Come correttivo per evitare abusi nell'uso di questi strumenti, la Corte ribadisce che questi strumenti hanno un ruolo meramente strumentale, funzionale ad individuare le esigenze specifiche dell'imputato.

Dai casi presentati risulta evidente come, di fatto, questi strumenti finiscano per avere un ruolo rilevante in fase di commisurazione della pena, poiché influenzano in maniera massiccia le conclusioni dei giudici, finendo per sostituirsi alle loro valutazioni critiche. Il rischio maggiore è che il COMPAS condizioni le decisioni dei giudici, sostanziandosi nell'elemento che più di ogni altro influisce nella determinazione della pena, come sostanzialmente accaduto nel caso Loomis.

La volontà di non commettere errori in tribunale si traduce così, il più delle volte, in una palese violazione dei diritti fondamentali degli imputati, giudicati non tanto per quanto hanno fatto ma per quanto potenzialmente *potrebbero fare* in futuro. La Corte del Wisconsin non sembra realizzare quanto in realtà questi strumenti influiscano sulle determinazioni finali dei giudici, in una società in cui le persone sono fortemente influenzate dall'«*automation bias»* che porta a riporre un'eccessiva fiducia verso questi strumenti[373]: «*automation bias effectively turns a computer program's suggested answer into a trusted final decision*[374].» Inoltre «*it ignores judges' inability to evaluate risk assessment tools, and it fails to consider the internal and external pressures on judges to use such assessments*.»

[373] K. Freeman, *Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State v. Loomis*, 18 N.C. J.L. & Tech. 75 (2016), p.98

[374] D. Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1271 (2008)

Nel decidere sulla detenzione o sull'adozione di altri provvedimenti cautelari non ci si può accontentare di una mera percentuale di rischio[375], calcolata tra l'altro sulla base di dati di gruppo (non collegati alla specifica criminale) e fattori per lo più statici, cioè indipendenti dal soggetto. Il fatto di vivere in un quartiere con un elevato tasso di criminalità o avere un livello di istruzione minore non implica certamente l'aver commesso un crimine o il poterlo compiere in futuro. Come osserva Nieva-Fenoll «finisce per essere più comodo giudicare la persona da quei fattori esterni, facilmente determinabili, che dalla vera responsabilità, che è molto più complessa da accertare[376] ».

Si potrebbe ritagliare un uso legittimo di questi programmi in fase di applicazione della sentenza, in un momento in cui la responsabilità dell'imputato è accertata in relazione a uno specifico fatto, ma usarli nelle precedenti fasi determina la sostituzione della decisione del giudice a un criterio matematico imperfetto e discriminatorio[377].

Sarebbe auspicabile che l'intelligenza artificiale fungesse da correttivo agli errori umani: bisognerebbe usare i contributi di questa scienza «nell'indagine dei fatti e non nella scoperta di circostanze esterne che nulla hanno a che fare con la responsabilità di un atto criminale[378].»

[375] J. Nieva- Fenoll, *op. cit.*, p. 62
[376] *Supra* nota 373, p. 65
[377] *Ibidem*
[378] *Ibidem*

CAPITOLO V

IMPATTO DEGLI ALGORITMI SUL DIRITTO ALL'EQUO PROCESSO

5.1 L'equo processo

Il diritto all'equo processo costituisce l'architrave di ogni ordinamento fondato sul principio di preminenza del diritto: responsabilità di ogni Stato democratico è assicurare un sistema di giustizia in cui il ricorrente possa riporre fiducia, formalmente e sostanzialmente equo.

Il principio è espressamente riconosciuto a livello universale in numerosi strumenti internazionali, tra cui la Dichiarazione Universale dei diritti dell'uomo (art. 10[379]) e il Patto Internazionale sui diritti civili e politici del 1966 (art. 14); a livello regionale è invece riconosciuto nella Carta dei diritti fondamentali dell'Unione Europea (art. 47), oltre che nella Convenzione europea per la salvaguardia dei diritti dell'uomo e delle libertà fondamentali (da qui CEDU) all'art.6. Oggetto di approfondimento della trattazione saranno proprio i principi contenuti in quest'ultima, in virtù del loro ampio contenuto dispositivo.

Le garanzie enunciate si applicano sia alle controversie civili che penali e comprendono il diritto ad un'equa e pubblica udienza[380], il diritto a essere giudicato entro un termine ragionevole[381], l'indipendenza e imparzialità del giudice[382], il

[379] «Ogni individuo ha diritto, in posizione di piena uguaglianza, ad una equa e pubblica udienza davanti ad un tribunale indipendente e imparziale, al fine della determinazione dei suoi diritti e dei suoi doveri, nonché della fondatezza di ogni accusa penale che gli venga rivolta.»

[380] Il diritto in questione può essere compresso in presenza di ragione di sicurezza, moralità, ordine pubblico o tutela della vita privata delle parti.

[381] Si intende per tale anche la decisione finale e la fase di esecuzione della sentenza ed è strumentale a un efficace ricorso al giudice.

[382] Il giudice deve essere indipendente rispetto all'esecutivo e al potere legislativo e equidistante rispetto alle parti a livello formale e sostanziale; non solo, dunque, non deve

diritto a un contraddittorio effettivo, il principio della parità delle armi e l'adeguata motivazione della sentenza[383]. Tra i diritti impliciti vi è quello di accesso *effettivo* ad un tribunale per ottenere una decisione utile alla definizione della domanda giudiziale. Ciò implica che lo Stato non può frapporre ostacoli sostanziali (ad esempio un onere economico eccessivamente gravoso) e procedurali (fissazione di termini particolarmente brevi per la proposizione del ricorso) all'esercizio del proprio diritto di azione[384].

Il primo paragrafo dell'art. 6 annuncia che «Ogni persona ha diritto a che la sua causa sia esaminata equamente, pubblicamente ed entro un termine ragionevole da un tribunale indipendente e imparziale, costituito per legge, il quale sia chiamato a pronunciarsi sulle controversie sui suoi diritti e doveri di carattere civile o sulla fondatezza di ogni accusa penale formulata nei suoi confronti.»

La presunzione di innocenza trova puntuale formulazione al comma 2: «Ogni persona accusata di un reato è presunta innocente fino a quando la sua colpevolezza non sia stata legalmente accertata.»

Al paragrafo terzo sono invece elencati una serie di requisiti specifici del processo penale:

a) «Essere informato, nel più breve tempo possibile, in una lingua a lui comprensibile e in modo dettagliato, della natura e dei motivi dell'accusa formulata a suo carico». Tale enunciato è funzionale all'esercizio del diritto di difesa dell'accusato, proprio perché la comprensione del fatto contestato, l'oggetto del processo, la data e luogo del fatto e la qualificazione giuridica sono necessari alla formulazione di una adeguata difesa.
b) «disporre del tempo e delle facilitazioni necessarie a preparare la sua difesa». L'efficacia della difesa è, di fatto, condizionata da numerosi fattori (tra cui il tempo) da valutare in relazione a tutte le peculiarità del caso concreto.

---

nutrire alcun interesse personale o altrui nella definizione della controversia, ma deve apparire anche come tale.

[383] Funzionale ad assicurare la trasparenza della giustizia e la comprensibilità del processo logico e decisionale seguito dal giudice nella definizione della controversia.

[384] P. Pustorino, *Lezioni di tutela internazionale dei diritti umani*, Cacucci Editore, Bari,2019, p. 147 e ss.

c) «difendersi personalmente o avere l'assistenza di un difensore di sua scelta e, se non ha i mezzi per retribuire un difensore, poter essere assistito gratuitamente da un avvocato d'ufficio, quando lo esigono gli interessi della giustizia».

d) «esaminare o far esaminare i testimoni a carico e ottenere la convocazione e l'esame dei testimoni a discarico nelle stesse condizioni dei testimoni a carico».

Un contraddittorio efficace richiede la possibilità di contestare una testimonianza a proprio carico interrogandone l'autore; se la sentenza si basasse unicamente su dichiarazioni che l'accusato non ha mai avuto modo di interrogare o far interrogare, si profilerebbe una violazione del diritto di difesa.

e) «farsi assistere gratuitamente da un interprete se non comprende o non parla la lingua usata in udienza».

Il presente capitolo si propone di esplorare le conseguenze lesive che l'ingresso degli algoritmi nel processo penale comporterebbero per il diritto all'equo processo, così come elaborato dalla giurisprudenza della Corte EDU. Ipotizzando future prospettive di apertura all'ingresso degli algoritmi nel sistema giudiziario, si tenterà nei prossimi paragrafi di fornire al lettore degli strumenti di analisi adeguati, per rilevare le criticità implicate rispetto alle garanzie dell'art. 6. Si procederà con ordine, al fine di verificare quanti e quali ordini di problemi sono implicati, anche surrettiziamente, nell'impiego degli strumenti predittivi in ambito processuale penale.

Una prima questione riguarda il vizio di opacità dell'algoritmo, che impedisce di disporre delle facilitazioni necessarie a predisporre la propria difesa. Nel prosieguo del lavoro si esporranno i rischi connessi alla parità delle armi e, soprattutto, i problemi legati alla massiccia influenza che tali strumenti esercitano sulle determinazioni finali dei giudici, finendo con assumere quasi il valore di una vera e propria *prova* non contestabile dalle parti in giudizio.

Come si delineerà più approfonditamente in seguito, il diritto di esaminare il testimone dovrebbe ricomprendere il diritto di esaminare le regole sottostanti al

risk-scoring methodology, dunque non solo il modello di calcolo e il suo funzionamento generale, ma anche lo scrutinio dei dati inseriti in esso.

Emergono chiaramente le problematicità sottese se si pensa che i risultati forniti dagli algoritmi non è affatto, come si dimostrerà, un risultato *neutrale*: la qualità dei dati inseriti e gli eventuali *bias* incorporati in essi influiscono sul risultato finale, con pregiudizio al divieto di discriminazione enunciato all'art. 14 e alla presunzione di innocenza.

Le prospettive non sono incoraggianti in tal senso: l'uso "compulsivo" degli strumenti di automazione in ogni ambito del quotidiano lascia presagire che in un futuro non troppo remoto gli algoritmi inizieranno a muovere i primi passi anche nella direzione del processo penale. Assumere consapevolezza in anticipo sui rischi che tale passo può comportare è utile per evitare un impiego non solo "inconsapevole" e passivo, ma anche lesivo delle libertà fondamentali dei soggetti coinvolti nei processi.

5.2 Il diritto di accesso all'algoritmo

Presupposto fondamentale ai fini dell'equità del processo è che l'imputato sia messo in condizione di preparare adeguatamente la sua strategia difensiva: egli potrà tutelarsi in modo pieno ed effettivo solo se sarà messo nelle condizioni di accedere a tutti gli elementi idonei a fondare la decisione giudiziale.

In ottica processuale assume rilevanza il diritto di accedere all'algoritmo[385], inteso come diritto di accedere alla logica e al funzionamento specifico dello stesso: il risultato fornito dall'algoritmo è infatti assimilabile a una *perizia* e, in quanto tale,

[385] «Tale conoscibilità dell'algoritmo deve essere garantita in tutti gli aspetti: dai suoi autori al procedimento usato per la sua elaborazione, al meccanismo di decisione, comprensivo delle priorità assegnate nella procedura valutativa e decisionale e dei dati selezionati come rilevanti. Ciò al fine di poter verificare che gli esiti del procedimento robotizzato siano conformi alle prescrizioni e alle finalità stabilite dalla legge o dalla stessa amministrazione a monte di tale procedimento e affinché siano chiare – e conseguentemente sindacabili – le modalità e le regole in base alle quali esso è stato impostato». Sul punto Cons. Stato, sez. VI, 8 aprile 2019, n. 2270

inquadrabile tra i mezzi di prova processuali, idonei a fornire elementi da porre a fondamento della decisione. A causa del vizio di opacità [386] , il *software* dell'algoritmo non è scrutinabile: ciò comporta l'utilizzo in giudizio di un risultato il cui procedimento generativo non è noto né accessibile, dunque inquadrabile come *ostensione del materiale probatorio* necessario alla preparazione della difesa.

Secondo la Corte, «la mancata ostensione di materiale probatorio, comprendente elementi che avrebbero potuto portare all'assoluzione oppure ad una riduzione della pena, può costituire diniego delle facilitazioni necessarie per la preparazione della difesa, e dunque una violazione dell'art. 6 § 3 lett. b) della Convenzione[387].»

Si consideri, inoltre, che le garanzie di accesso al mezzo di prova sono ormai state riconosciute dalla Corte per qualsiasi tipologia probatoria, compresi *computer files* rilevanti per le accuse mosse all'imputato [388]: nel caso *Georgios Papageorgiou v Greece* la Corte ha riscontrato una violazione del diritto all'equo processo del ricorrente proprio perché era stata rifiutata la richiesta di produzione di estratti di *file* di un computer (funzionali alla difesa), ritenuta superflua dalla corte d'Appello di Atene.

Ecco, dunque, che il diritto di accesso all'algoritmo può essere ricondotto nell'alveo delle garanzie predisposte dalla CEDU in ambito di diritto a disporre delle facilitazioni necessarie a preparare la propria difesa. Come rilevato nel caso *Gregačević v. Croatia[389]*, «*The accused must have the opportunity to organise his defence in an appropriate way and without restriction as to the ability to put all relevant defence arguments before the trial court and thus to influence the outcome of the proceedings*.[390]»

---

[386] *Infra* para. 2.2

[387] Eur. Court of human rights , 4th Section, 31.3.2009, 21022/04, *Natunen v. Finland,* , § 43 «. Failure to disclose to the defence material evidence, which contains such particulars which could enable the accused to exonerate himself or have his sentence reduced would constitute a refusal of facilities necessary for the preparation of the defence, and therefore a violation of the right guaranteed in Article 6 § 3 (b) of the Convention»

[388] Eur. Court of human rights 1st Section, 9.5.2003, 59506/00, *Georgios Papageorgiou v Greece* , §37.

[389] Eur. Court of human rights, 1st Section ,10. 7.2012, 58331/09*, Gregačević v. Croatia,* § 51

[390] Sul punto anche Mayzit v. Russia, no. 63378/00, § 78, 20 January 2005; Connolly v. the United Kingdom (dec.), no. 27245/95, 26 June 1996; Can v. Austria, no. 9300/81

L'accusato non deve, dunque, essere ostacolato nell'ottenere copia di tutti i documenti rilevanti del giudizio: è possibile ricondurre nel perimetro di tale garanzia anche la possibilità di accedere al codice sorgente e alle specifiche del *software* algoritmico.[391]

Al netto di tali considerazioni, appare evidente come l'opacità del procedimento decisionale algoritmico mini *in nuce* il diritto di difesa, che implica di poter svolgere in modo effettivo tutte le argomentazioni e le prove a favore[392]. Alle parti è così preclusa la possibilità di verificare l'accuratezza dei dati e contestare il risultato algoritmico in maniera efficace, non potendo conoscere i passaggi che hanno determinato un certo risultato, la qualità dei dati inseriti ed il peso a loro attribuito, il codice sorgente e le sue specificità tecniche.

A tal proposito, come si è osservato nel caso Loomis, la Corte del Wisconsin ha negato il problema dell'accessibilità all'algoritmo, sostenendo che il risultato del COMPAS non fosse l'unico elemento su cui il giudice aveva fondato la decisione: la risposta risulta inadeguata perché consente l'innesto di un elemento di prova non scrutinabile dalle parti[393].

La mancanza di informazioni significative sulla logica algoritmica non consente di determinare la rilevanza data a specifici dati che possono avere un impatto discriminatorio sulle minoranze, i criteri di selezione e classificazione utilizzati, la presenza di eventuali *bias* nei dati immessi nell'algoritmo per creare i modelli statistici e in che modo evolve l'algoritmo con quei dati[394].

Tanto più risultano preoccupanti tali conclusioni se si pensa che gli algoritmi sono utilizzati in ambiti delicati, che comportano effetti giuridici sulla sfera di libertà dei soggetti coinvolti e non è ammissibile che i tribunali si trincerino dietro la difficoltà tecnica del funzionamento del programma[395] per eludere il problema.

---

[391] Rasmussen v. Poland, §§ 48-49; Moiseyev v. Russia, §§ 213-218; Matyjek v. Poland, § 59; Seleznev v. Russia, §§ 64-69
[392] J. Nieva- Fenoll, *op. cit*., p. 129
[393] J. N. Fenoll, *op. cit*., p. 132
[394]N. Diakopoulos, *Algorithmic Accountability Reporting: On the Investigation of Black Boxes* in *Tow Center for Digital Journalism*, Columbia University, July 10 2014
[395] J. Nieva-Fenoll, *op.cit.*, p. 63

Il rischio è che il giudice, per comodità, si affidi a un criterio di cui non conosce il funzionamento e che è intrinsecamente imperfetto e spesso discriminatorio, delegando ad esso decisioni di esclusiva competenza umana.

Nei sottoparagrafi che seguono verranno analizzate le ragioni ostative all'accesso all'algoritmo e il difetto di pubblicità e trasparenza che ne consegue.

5.2.1 Il difetto di pubblicità e trasparenza del meccanismo decisionale algoritmico

Con riguardo agli strumenti algoritmici, emerge il complesso rapporto tra tutela della proprietà intellettuale ed esigenza di trasparenza dei processi computazionali coinvolti nella decisione. Sempre più frequentemente le compagnie private, per rafforzare la loro competitività sul mercato, negano l'accesso alle specifiche tecniche dei *software* algoritmici, opponendo il segreto industriale. Questo determina l'utilizzo nel processo di programmi di cui i litiganti (e il giudice) non conosco i dettagli funzionali, senza contare che l'intelligenza artificiale conduce l'attività difensiva su un piano molto tecnico. Infatti, le parti dovranno saper leggere le specifiche tecniche degli algoritmi e il loro funzionamento per poter preparare una difesa adeguata.

Emerge in maniera evidente come la tutela della proprietà intellettuale dell'algoritmo confligga con le norme a presidio delle garanzie processuali, in particolare quelle volte a garantire un efficace contraddittorio.

L'esigenza di trasparenza del processo decisionale algoritmico è stata fortemente sentita a livello internazionale: numerosi documenti e pronunce giudiziali in materia (in particolare articoli 13 e 15 del GDPR) fissano il diritto di accesso all'algoritmo, inteso come possibilità di accedere non solo al meccanismo di funzionamento dello stesso, ma anche alla qualità dei dati immessi e al peso attribuito a ciascuna delle variabili.

Come richiamato anche all'interno della Carta Etica «deve essere raggiunto un equilibrio tra la proprietà intellettuale di alcune metodologie di trattamento e l'esigenza di trasparenza (…) quando si utilizzano strumenti che possono avere

conseguenze giuridiche, o che possono incidere significativamente sulla vita delle persone.[396] »

Il principio di *trasparenza algoritmica* può essere qualificato in termini generali come «l'obbligo, gravante sui soggetti che adottano decisioni con l'ausilio di sistemi automatizzati di trattamento dei dati, di fornire ai destinatari una spiegazione comprensibile delle procedure utilizzate e di motivare sotto questo profilo le decisioni assunte.[397]»

Per garantire a pieno la decodifica di tali strumenti, le informazioni devono essere non solo accessibili, ma anche *comprensibili*. Come osservato sul punto «*transparency is not enough, in itself: transparency must be meaningful; the disclosure of the source code is not considered true transparency, because only experts can understand it.*[398]»

Lo stesso art. 15 del GPDR segna all'interno del principio di trasparenza la necessità che sia utilizzato un linguaggio semplice e comprensibile. Il principio di comprensibilità si collega così al diritto di accesso all'algoritmo, cioè la garanzia di un accesso quanto più ampio possibile alle informazioni algoritimiche, potendo così individuare sia «i suoi autori, sia il procedimento usato per la sua elaborazione, sia il meccanismo di decisione, comprensivo delle priorità assegnate nella procedura valutativa e decisionale e dei dati selezionati come rilevanti»[399].

L'accesso alle specifiche dell'algoritmo e al codice sorgente allo stato attuale risulta improbabile a causa delle politiche delle società private a tutela del segreto industriale. La soluzione potrebbe essere quella di rendere accessibile una certa quantità di informazioni predeterminate, ad esempio quali variabili sono utilizzate, per quale obiettivo è stato ottimizzato l'algoritmo, la tipologia e quantità di dati immessi, il modo in cui vengono monitorate le prestazioni dell'algoritmo, come l'algoritmo stesso evolve nel tempo, i fattori rilevanti per il funzionamento

[396] CEPEJ, *Carta Etica,* p.11
[397] P. Zuddas, *Brevi note sulla trasparenza algoritmica* in *Amministrazione in Cammino*, 5 giugno 2020
[398] S. Quattrocolo, C. Anglano, M. Canonico, M. Guazzone, *Technical Solutions for Legal Challenges: Equality of Arms in Criminal Proceedings* in *Global Jurist,* 2020
[399] *Ibidem*

dell'algoritmo, i dati inseriti per il suo "addestramento", la loro classificazione e il peso attribuito a ciascuno di essi[400].

L'importante non è rendere accessibili tutte le informazioni, ma quelle più «significative e funzionali per assicurare una trasparenza effettiva»[401].

5.2.2 L'inaccessibilità del codice sorgente e il vizio di opacità dell'algoritmo

L' opacità tecnica indica la situazione tale per cui il risultato dell'algoritmo o non è *conoscibile* o non risulta *comprensibile*.

In riferimento a modelli di cui è possibile descrivere solo il comportamento esterno (*output*), ma di cui non si è in grado di conoscere il funzionamento interno ricostruendo il percorso logico che ha condotto al risultato, è stata coniata l'espressione "*black boxes*"[402]. In questo caso «*the input and ultimate output of the system are observable, but how the system arrives at that outcome is unknown, even to those who created it*[403].»

Il crescente affidamento fatto sui *Big Data* per prendere qualsivoglia tipo di decisione ha inasprito il problema, al punto che per descrivere la pervasività del fenomeno si parla di *«black box society»*[404].

Sussistono tre circostanze che possono determinare tale vizio[405]:

---

[400] L. McGregor., D. Murray, & Ng,V., *International human rights law as a framework for algorithmic accountability* in *International and Comparative Law Quarterly,*Vol. 68, 2019, p. 309-343.

[401] Council of Europe, *Algorithms and human rights - Study on the human rights dimensions of automated data processing techniques and possible regulatory implications* (2018)

[402] S. Quattrocolo, *Processo penale e rivoluzione digitale: da ossimoro a endiadi?* In *Medialaws*, 3/2020

[403] L. Tiller, *A Minority Report: The Unregulated Business of Automating the Criminal Justice System* in The Business, Entrepreneurship & Tax Law Review's B.E.T.R. White Paper, Marzo 2019, p. 10 e ss.

[404]Frank Pasquale, *The black box society: The secret algorithms that control money and information*, Harvard University Press (2015)

[405] Burrell J. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*. June 2016, p. 1 e ss.

i. L'opacità come obiettivo intenzionalmente perseguito delle politiche di tutela del segreto industriale adottate dalle compagnie private, che rendono inaccessibili le specifiche tecniche dell'algoritmo e il codice sorgente[406] per mantenere dei vantaggi competitivi rispetto ai concorrenti sul mercato. Molto spesso si parla di tale vizio come forma di «*proprietary protection*» o come «c*orporate secrecy*»[407].

ii. Opacità dovuta alle competenze tecniche richieste per l'intelligibilità del risultato, non alla portata dei comuni cittadini. Inoltre, se il design dell'algoritmo non è improntato alla trasparenza, potrebbe essere impossibile verificare l'attendibilità dell'output per chiunque non sia il *designer* del codice sorgente stesso[408].

iii. Opacità intrinseca ai sistemi di *machine learning*: tali sistemi operano secondo una logica "deduttiva", evolvendosi e imparando da dati di volta in volta immessi. La conseguenza è che anche qualora si rivelasse il codice sorgente, potrebbero comunque risultare non pienamente comprensibili le ragioni e i passaggi seguiti dalla macchina per fornire quel determinato risultato (si parla di opacità *intrinseca*).

Il problema, come già evidenziato, è fortemente sentito nel sistema giudiziario statunitense: secondo i giudici è ammesso l'utilizzo degli algoritmi in fase di *sentencing*, in quanto la possibilità di contestare il risultato facendo ricorso al manuale d'uso dell'algoritmo è sufficiente a garantire il diritto di difesa[409]. Tale argomento non persuade se si tiene conto, come osservato, «del peso psicologico che può esercitare sul giudicante il risultato di un algoritmo che non dà conto dei

[406] Scrittura dell'algoritmo in uno specifico linguaggio di programmazione, ad es. Python
[407] *Ibidem*
[408] S. Quattrocolo, *Equità del processo penale e automated evidence alla luce della Convenzione europea dei diritti dell'uomo* in *Revista Ítalo-Española de Derecho Procesal*, Vol. 1 | 2019
[409] C. Cesari, *Editoriale: L'impatto delle nuove tecnologie sulla giustizia penale – un orizzonte denso di incognite* in *Revista brasileira de direito processual penal, Porto Alegre, vol. 4, n. 3,* p. 1177 e ss.

dati, ma li ricostruisce per vie misteriose in un "pacchetto decisorio" preconfezionato con una soluzione data.[410]»

Il diritto di difesa implica una serie di garanzie, tra cui la possibilità di verifica (in termini di piena accessibilità e comprensibilità) di qualunque contributo utile a influenzare la decisione giudiziale. Il vizio di opacità determinato dalla segretezza del *software* preclude la verifica dei risultati, che dovrebbe sempre essere sempre assicurata[411].

È evidente che le delicate fasi del processo non possono essere regolate da una "scatola nera", non sono ammissibili spazi "oscuri" alle parti in quanto «se non fosse possibile nemmeno conoscere questo contenuto per ragioni di proprietà intellettuale, allora il diritto di difesa cesserebbe di esistere[412].»

Prendendo nuovamente in esame il caso dei *risk assessment tool,* come ampiamente illustrato[413], è ipotesi affatto rara che i dati inseriti nell'algoritmo siano viziati da pregiudizi razziali: a causa dell'effetto *black box* l'accusato o lo stesso giudice non avranno contezza di tale vizio.

La soluzione, dunque, va ricercata nello sforzo di elaborare regole chiare, precise e un nucleo duro di informazioni che, al di là del segreto industriale, devono essere estrinsecate per assicurare l'equità del processo.

Nella prima forma di opacità, la soluzione sarebbe data dal rendere disponibile il codice sorgente (c.d. *open source code*) dell'algoritmo in modo tale da poter individuare e correggere eventuali manipolazioni o errori[414]. Si è osservato tuttavia che questo non è un rimedio effettivo, in quanto solo esperti del settore sarebbero in grado di comprendere il significato del codice. La piena comprensibilità è assicurata a condizione che sia corredata da *spiegazioni* che la traducano nella

[410] *Ibidem*
[411] *Ibidem*
[412] J.Nieva-Fenoll, op.cit.
[413] *Supra* Capitolo IV,para. 7.3
[414] La Carta etica suggerisce ad esempio come rimedio la creazione di autorità preposte alla verifica e certificazione dei modelli automatici impiegati nel processo.

«regola giuridica» ad essa sottesa e che la rendano *leggibile* e *comprensibile*, sia per i cittadini che per il giudice[415].

Altra soluzione potrebbe essere il ricorso a un perito indipendente che *ex post* possa valutare l'attendibilità del risultato algoritmico[416]. Tra le altre soluzioni prospettate si è avanzata l'ipotesi di sviluppare dei *software* di crittografia detti "*zero-knowledge proof*" con cui sarebbe possibile individuare i criteri che governano la *policy* dell'algoritmo, senza dover svelare la policy stessa[417] per poter verificare la correttezza dell'*output*. Tale *software*, tuttavia, non è utile per valutare l'affidabilità dei dati raccolti per mezzo di una catena algoritmica[418].

5.3 La lesione del diritto alla parità delle armi

Pur non essendo esplicitamente menzionato dall'art. 6 della CEDU, la parità delle armi costituisce l'essenza dell'equità processuale, in quanto strumento di garanzia di un "*fair balance"* tra accusa e difesa. Il principio richiede che ad ogni parte sia data l'opportunità di difendere le proprie ragioni in condizioni di parità, nonché poteri equivalenti[419], evitando che una si trovi in condizioni di svantaggio rispetto alla controparte[420].

Presupposto del principio di parità delle armi è il contraddittorio nella formazione della prova: deve essere garantita alle parti la possibilità di aver cognizione delle prove e degli atti dedotti nel processo e poter contro-dedurre in ordine ad essi, così da poter influenzare a proprio favore la decisione del

[415] Consiglio di Stato, sez. VI, 8 aprile 2019, n. 2270
[416] *Supra* nota 382, p.121 e ss.
[417] *Ibidem*
[418] S.Quattrocolo, C. Anglano, M. Canonico, M.Guazzone, *Technical Solutions for Legal Challenges: Equality of Arms in Criminal Proceedings* in *Global Jurist,* 2020, p. 15 e ss.
[419] G. Illuminati, *Giudizio* in *Compendio di procedura penale* a cura di G. Conso, V. Grevi, Wolters Kluewer, Milano, 2018 p. 742
[420] Eur. Court of human rights, Grand Chamber, 7.6.2001, 39594/98, K*ress v*. *France*. Sul punto anche *Foucher v. France,* § 34; *Bulut v. Austria; Bobek v. Poland,* § 56; *Klimentyev v. Russia*, § 95

tribunale[421]. Sono escluse da tale concezione le letture riduttive del principio dialettico della formazione dibattimentale della prova, secondo le quali sarebbe sufficiente la discussione in giudizio sui risultati delle prove acquisite altrove: l'unica prova utilizzabile è quella formata davanti al giudice con intervento delle parti[422].

Il principio risulta violato laddove "*is denied the opportunity to attend the proceedings, or where he is unable properly to instruct his legal representative*[423]".

Nel caso *Kuopila v. Finlanda* la Corte ha statuito che la mancata ostensione delle prove alla difesa costituisce una violazione del principio alla parità delle armi. Alla difesa era stato precluso di contro-dedurre rispetto ad un'integrazione di informativa della polizia giudiziaria[424]. Sul punto la Corte ha osservato che «*the procedure did not enable the applicant to participate properly and in conformity with the principle of equality of arms in the proceedings before the Court of Appeal*.»

Condizione essenziale per il rispetto di tale principio è l'accessibilità, la conoscenza degli elementi di prova e degli argomenti dedotti dalle controparti in giudizio: l'opacità degli algoritmi impedisce tale dialettica, ammettendo surrettiziamente l'ingresso in giudizio di un contributo non controvertibile dalla difesa.

Ecco, dunque, che l'introduzione di prove algoritmiche nel processo comporta la potenziale violazione del *fair trial:* le prove (o più in generale gli elementi che

[421] Consiglio d'Europa / Corte Europea dei Diritti dell'Uomo, *Guida all'art. 6,* 2014, p. 23 e ss., consultabile all'indirizzo https://www.echr.coe.int/Documents/Guide_Art_6_criminal_ITA.pdf

[422] Ibidem (G. Illuminati) p 749

[423] Communication No. 289/1988, D. Wolf v. Panama (Views adopted on 26 March 1992), in UN doc. GAOR, A/47/40, pp. 289-290, para. 6.6

[424] Eur. Court of human rights, 4th Section, 27.4.2000, 27752/95*, Kuopila v. Finland,* § 38. Nello specifico «In the instant case, the prosecutor had expressed his opinion on the relevance of the report to the Court of Appeal, thereby intending to influence the court's judgment. The Court considers that procedural fairness required that the applicant too should have been given an opportunity to assess the relevance and weight of the supplementary police report and to formulate any such comment as she deemed appropriate. It is also noted that the applicant had requested a supplementary investigation and that throughout the proceedings she had considered it to be important.»

possono influenzare la commisurazione finale della pena) generati da *software* o sistemi computazionali, impediscono alla difesa di validarne la genesi e dunque la genuinità del dato[425]. Di fatto, l'argomento basato sull'*output* di un algoritmo non scrutinabile diventa un «argomento di cui può avvalersi una sola parte, poiché la difficoltà di spiegarne la genesi si trasforma in una difficoltà di contestarne l'attendibilità»[426].

L'impossibilità per la difesa di contestare l'accuratezza e, quindi, l'attendibilità della prova a carico, produce un forte squilibrio di potere tra accusa e difesa. La potenziale presenza di un *bias* discriminatorio nell'algoritmo determina di conseguenza un danno nei confronti della minoranza oggetto del pregiudizio, il più delle volte inconsapevole di subire l'ingiustizia a causa dell'effetto *black box.*[427]

5.3.1 L'asimmetria conoscitiva tra le parti in giudizio e il diritto di esaminare il testimone a carico

Lo strumento algoritmico, richiedendo particolari conoscenze tecniche per essere valutato, esaspera il fenomeno dell'asimmetria conoscitiva tra le parti nel processo (c.d *knowledge asymmetry*).

L'accuratezza e l'attendibilità dei dati utilizzati dall'algoritmo non è verificabile a causa del vizio di *opacità* del modello. Ragion per cui, viene meno l'effettiva possibilità di contestare il risultato fornito per la difesa, in posizione di svantaggio rispetto all'accusa che ha accesso a risorse tecnologiche i cui risultati sono trasferiti nel processo penale. Questo implica che la difesa dell'accusato subirebbe una restrizione incompatibile con l'equità del processo se una sentenza di condanna si basasse su «dichiarazioni che l'accusato non ha mai avuto modo di interrogare o fare interrogare.[428]»

---

[425] S. Quattrocolo*, Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale "predittiva"* in *Cassazione penale* n .4/2019, p.1761 e ss.
[426] *Ibidem*
[427] *Supra* nota 391
[428] V. Zagrebelsky, *Manuale dei diritti fondamentali in Europa*, Il Mulino, Bologna, 2019, p. 241

Sicchè, il dato ottenuto automaticamente rischia di divenire attendibile *di per sè,* perché la verifica del processo che lo ha generato è troppo complessa o sfugge, almeno in parte, ad un controllo *ex post* [429]. Anche il giudice è portato ad adagiarsi sul risultato fornito, non avendo la difesa potuto addurre elementi convincenti per discostarsene.

Ecco, quindi, che l'algoritmo «introduce la forma più estrema di tale squilibrio, poiché il risultato probatorio può essere non criticabile laddove, appunto, l'inaccessibilità del codice sorgente o altre caratteristiche del software non consentano alla parte contro la quale la prova è introdotta nel processo di contestarne l'accuratezza e l'attendibilità.»[430]

Lo squilibrio tra parte pubblica e difesa è connaturato alla struttura del processo penale, soprattuto nella fase preliminari al giudizio. Ad ogni modo, in tali casi la parità delle armi è garantita dal fatto che a tutti è data la possibilità di presentare i propri argomenti in condizioni che non la svantaggino rispetto alle altre[431].

A tal proposito, nel caso *Brandstetter c. Austria*[432] la Corte ha ribadito che è necessario che ciascuna parte abbia *effettiva* [433] (e non meramente teorica) conoscenza delle allegazioni e delle argomentazioni della controparte e che fruisca della concreta possibilità di contestarle [434] poichè *«an indirect and purely hypothetical possibility for an accused to comment on prosecution argument*[435]*»*. Se le caratteristiche interne dell'elemento di prova si basano esclusivamente su un processo computazionale, lo spazio per la critica è ostacolato.

---

[429] *Supra* nota 382
[430] *Supra* nota 382
[431] *Ibidem*
[432] Eur. Court of human rights , Chamber, 28.8.1991, 111170/84; 12876/87; 13468/87, *Brandstetter c. Austria*
[433] Al fine di verificare se il principio sia stato rispettato, la Corte in *Bykov v. Russia* sostiene come occorra «verificare in particolare se il richiedente è stata data l'opportunità di contestare l'autenticità delle prove e di opporsi il suo utilizzo. Inoltre, la qualità delle prove deve essere presa in considerazione, compreso se le circostanze in cui sono state ottenute mettono in dubbio la sua affidabilità o accuratezza.». Eur. Court of human rights, Grand Chamber, 10.3.2009, 4378/02, *Bykov v Russian Federation*
[434] «the parties must be aware of the opponent's statements and allegations and get "a real opportunity to comment" on it» (Brandstetter v. Austria, cit., § Par. 67)
[435] *Supra* nota 432, § 68

Emerge chiaramente la lesione della parità delle armi nell'impiego di contributi automatizzati in fase processuale. Infatti, la difesa non disporrà di tutti gli elementi necessari per poter contrastare e valutare l'accuratezza del risultato fornito, che pur non essendo l'unico elemento su cui si fonda la decisione, influenza profondamente le determinazioni del giudice[436]. Così «l'impossibilità di verificare a posteriori l'output di un algoritmo può costituire in nuce una violazione dell'art. 6§1 Cedu[437].»

Ulteriore corollario dell'equo processo è il diritto di esaminare il testimone a carico e ottenere la convocazione e l'esame dei testimoni a discarico, volto ad assicurare che l'accusato abbia l'effettiva possibilità di confrontarsi con il testimone per rispettare il principio di produzione della prova in fase dibattimentale.

A dispetto di quanto suggerisce la nozione in prima lettura, il diritto non si applica solo all'esame dei testimoni. Di fatto, la nozione di testimone elaborata dalla giurisprudenza della corte è autonoma e particolarmente ampia, tale da ricomprendere tutti i contribuiti idonei quali basi per una condanna[438] e include (oltre ai coimputati e le persone offese) anche i periti ed i consulenti, cui è sostanzialmente assimilabile il contributo algoritmico[439]. Di conseguenza, si è osservato che «*in order to ensure effective participation in a trial, the defendant must also be able to challenge the algorithmic score that is the basis of his or her conviction.*[440]»

Sul punto è innovativa la pronuncia[441] con cui la Corte EDU afferma che l'assenza ingiustificata del testimone d'accusa il diritto al confronto dell'accusato,

---

[436] *Infra* para. 4 "L'effetto ancoraggio" per una analisi più approfondita sul tema.
[437] *Supra* nota 382, p. 120
[438] Eur. Court of human rights, 1st Section, 09/11/2006, 18885/04, *Kaste and Mathisen v. Norway,* § 53 Secondo cui «t should be reiterated that where a deposition may serve to a material degree as the basis for a conviction then, irrespective of whether it was made by a witness in the strict sense or by a co-accused, it constitutes evidence for the prosecution to which the guarantees provided by Article 6 §§ 1 and 3 (d) of the Convention apply.»
[439] Eur. Court of human rights, 3rd Section, 26.3.1996, 10524/92*, Doorson v Netherlands,* para. 81 dove si qualifica come testimone un esperto tecnico.
[440] A. Završnik, *Criminal justice, artificial intelligence systems, and human rights* in *ERA Forum,* 2020**,** p.577
[441] Eur. Court of human rights, 3rd Section, 10.2. 2015, 26504/06, *Colac v. Romania* , par. 39

risultando irrilevante che le dichiarazioni dei testimoni assenti non abbiano costituito la prova unica o determinante su cui si basa la condanna.

In *Al-Khawaja and Tahery c United Kingdom*[442] la Corte ha rilevato come il summenzionato diritto debba tradursi non solo nella possibilità di verificare la credibilità, ma anche l'*affidabilità* e *accuratezza* delle argomentazioni a carico. Come premesso nei precedenti paragrafi, l'opacità dell'algoritmo determina l'impossibilità *in nuce* di effettuare tali verifiche, determinando un potenziale svantaggio per l'accusato.[443] I problemi rappresentati dall' AI sono simili a quelli posti dal testimone anonimo o assente, di per sé non contrari all'art. 6 ma ammessi solo come estrema *ratio* e in condizioni tali da assicurare che l'accusato non versi in una posizione di svantaggio. Simili restrizioni dovrebbero essere applicate anche all'impiego degli strumenti algoritmici nel processo, introducendo così un rimedio alla situazione di squilibrio che si produce tra le parti.

Inoltre, il diritto di *cross-examination* dovrebbe essere interpretato anche nel senso di garantire il diritto di esaminare i dati e le regole di funzionamento dei *Risk Assessment tool* [444].

Non avendo la difesa alcuna possibilità di contraddire quanto determinato dall'algoritmo si viene a creare una situazione di sostanziale svantaggio per l'imputato, proprio come nel caso in cui siano ammesse testimonianze anonime, che per definizione impediscono la dialettica dibattimentale.

[442] Eur. Court of human rights, Grand Chamber, 15.12.2011, 26766/05 and 22228/06, *Al-Khawaja and Tahery v. the United Kingdom*, para. 41

[443] Per certi versi, i problemi sollevati dai sistemi di intelligenza artificiale utilizzati in giudizio possono essere assimilati a quelli del testimone anonimo o della prova documentale riservata;il testimone anonimo o assente non è di per sé incompatibile con l'equo processo a patto che sia ammesso solo in condizioni stringenti e come *extrema ratio*.

[444] *Supra* nota 437

5.4 Il rischio di pressione indiretta sul giudice

Pur non costituendo l'elemento essenziale su cui si fonda la decisione, nel caso *State vs Loomis* è stata analizzata la portata determinante dell'algoritmo nella decisione in fase di commisurazione della pena.

Si è opportunamente osservato come l'algoritmo sia in grado di esercitare pressioni interne ed esterne rispetto al giudice che si appresta a formulare la decisione e come anche i *bias* cognitivi incoraggiano l'uso di tali strumenti, ritenuti oggettivamente obiettivi e neutrali in virtù del c.d *automation bias*.

Si consideri che un giudice, dovendo decidere in condizioni di incertezza, difficilmente non si lascerà influenzare dal risultato dell'algoritmo: conscio dell'alto rischio di recidiva, non si assumerà il rischio di ricorrere a strumenti alternativi alla detenzione, né tantomeno comminerà una pena detentiva troppo breve.

Un giudice al quale viene fornita una valutazione del rischio che pronostica un alto tasso di recidiva «potrebbe essere portato a irrogare una pena maggiore senza aver neanche la minima consapevolezza del ruolo avuto dall'"*anchoring*" nella decisione medesima[445]».

Un ulteriore rischio è che il giudice possa valutare come colpevole il soggetto non per il fatto commesso, ma per la *futura probabilità* di commettere reati in futuro, in virtù di una inversione argomentativa[446] o basare le sue valutazioni sul profilo di personalità dell'imputato tracciato dall'algoritmo, piuttosto che sui fatti concretamente commessi e oggetto di giudizio.

Di fatto, se i *Risk Assessment Tool* fossero infallibili e funzionassero al riparo da pregiudizi e dinamiche discriminatorie, il loro contributo nel processo sarebbe non solo utile, ma anche auspicabile. Tuttavia, allo stato attuale dell'arte, per tutti i

[445] L. Maldonato, *Algoritmi predittivi e discrezionalità del giudice: una nuova sfida per la giustizia penale* in *Diritto penale contemporaneo*, 2/2019, p. 410
[446] *Ibidem*

profili di criticità precedentemente analizzati (non per ultimo quello dell'opacità) risulta essere un'ingerenza "rischiosa" e fuorviante. Oltre all'*automation bias* sopracitato[447], giocano un ruolo primario in questo senso il c.d. "effetto ancoraggio" e il rischio di delega della decisione all'algoritmo, che verranno delineati nei prossimi sottoparagrafi.

5.4.1 L'effetto ancoraggio

Come evidenziato in un esperimento condotto nel 2001 da un team di giuristi americani[448], anche i giudici possono essere influenzati da una serie di *bias* cognitivi, cioè distorsioni in grado di alterare il processo decisionale del giudice[449].

Sono stati analizzati gli effetti del c.d. *ancoraggio,* fenomeno in base al quale un decisore umano attribuisce un certo peso a un dato tangibile e immediatamente disponibile, in modo potenzialmente lesivo per la decisione[450].

Riprendendo in esame l'utilizzo dei *risk assessment tool* in fase di giudizio, è evidente come la stima dell'algoritmo circa il rischio di recidiva del soggetto possa fungere da "ancora" per le determinazioni del giudice, che sarà portato (anche solo inconsciamente) a emanare una sentenza in linea con quanto stabilito dallo strumento di intelligenza artificiale. Se il rischio di recidiva calcolato risulta alto, il giudice sarà inevitabilmente condizionato da tale dato nella commisurazione della pena da infliggere all'imputato e portato ad irrogare una pena più severa.

Nello specifico, il punteggio COMPAS viene presentato come tre grafici a barre, ciascuno dei quali mostra al giudice della condanna un punteggio da uno a dieci.

---

[447] *Supra* Capitolo IV, para. 8

[448] In particolare, i Professori. Chris Guthrie e Jeffrey J. Rachlinski e il giudice Andrew J. Wistricliff.

[449] S. Arceiri, *Bias cognitivi e decisione del giudice: un'indagine sperimentale* in *Diritto Penale e Uomo,* 2/2019. Ad esempio, quando è necessario effettuare una stima numerica (ad esempio, il valore di mercato di una casa), le persone tendono a fare affidamento sul primo dato a disposizione (ad esempio, il prezzo di listino). La stima finale tende ad "ancorarsi" a quel valore iniziale

[450] M. E. Donohue, *A replacement for justitia's scale? Machine learning in sentencing* in *Harvard Journal of Law & Technology*, Volume 32, Number 2, 2019, p. 661 e ss.

Questa chiara misura quantitativa può prevalere sugli altri fattori qualitativi e le intuizioni umane dovranno essere forti per poter andare oltre tale dato[451]. L'effetto è stato documentato nel caso *State vs Loomis.* Il giudice nella motivazione della sentenza ha espressamente richiamato il punteggio di rischio elevato, calcolato dal COMPAS, per giustificare la dura pena comminata all'accusato.

Quanto sopra riportato conferma che le ancore hanno un effetto pervasivo da cui è difficile discostarsi e il pregiudizio è amplificato se si pensa che i risultati sono forniti sulla base di dati che spesso nascondono dei pregiudizi impliciti. In sostanza, l'algoritmo è uno strumento «dall'estrema persuasività e il suo manto di oggettività, che lo avvicina, nel suo aspetto esteriore, ad una prova scientifica, può condizionare fortemente il giudice al momento della decisione[452].»

5.4.2 La decisione "delegata" all'algoritmo

Il caso di Eric Loomis è emblematico delle numerose criticità sottese all'uso degli algoritmi predittivi in fase di di giudizio. Il problema principale è l'affidamento che i giudici ripongono nei confronti di un programma di cui non è noto il funzionamento, minando in radice il diritto di difesa dell'imputato, impossibilitato a conoscere le basi sulle quali viene fornito il risultato. E anche laddove il giudice fosse consapevole di tali rischi, l'*outcome* algoritmico avrebbe già esercitato un "effetto ancoraggio" sulle sue determinazioni, rendendo difficile per quest'ultimo discostarsi dalle conclusioni dell'algoritmo[453].

Il software predittivo è infatti «un comodo riparo per il giudice che, nascondendosi dietro lo score, potrebbe omettere di considerare tutte le peculiarità del caso e, come immediata conseguenza, omettere di motivare adeguatamente in ordine alla commisurazione della pena.[454]»

---

[451] *Ibidem*
[452] *Supra* nota 416
[453] Gerards J. *The fundamental rights challenges of algorithms* in *Netherlands Quarterly of Human Rights*. 2019;37(3):205-209.
[454] *Supra* nota 416

La Corte del Wisconsin ha creato un precedente "pericoloso", aprendo le porte all'uso diffuso del COMPAS nei tribunali e affermando che il diritto di accesso all'algoritmo non è riconosciuto all'imputato e che questi non viene leso in alcun modo nel suo diritto a un *fair trial.* Come correttivo per evitare abusi nell'uso di questi strumenti, la Corte ribadisce che questi strumenti hanno un ruolo meramente strumentale, funzionale a individuare le esigenze specifiche dell'imputato. Tuttavia, dai casi presentati nel precedente capitolo, risulta evidente come, di fatto, questi strumenti finiscano per avere un ruolo ben più rilevante in fase di commisurazione della pena, poiché influenzano in maniera massiccia le conclusioni dei giudici, finendo per sostituirsi alle loro valutazioni critiche. Il rischio maggiore è che il COMPAS si sostanzi nell'elemento che più di ogni altro influisce nella determinazione della pena. Ciò in quanto «finisce per essere più comodo giudicare la persona da quei fattori esterni, facilmente determinabili, che dalla vera responsabilità, che è molto più complessa da accertare[455]».

L'AI può certamente affiancare il giudizio per certi aspetti, ma l'aspirazione all'efficienza «non può comportare una passiva e totale delega alle tecnologie informatiche dell'esercizio dell'attività giurisdizionale. In realtà, la qualità della giustizia è indissolubilmente legata a doppio filo con la sensibilità, l'esperienza e la capacità del giudice -persona fisica- di cogliere le piccole circostanze che rendono ogni decisione unica e non gestibile in modo standardizzato e statistico»[456].

Anche se l'ipotesi di un giudice-robot in sostituzione di un giudice umano sembra ancora un'ipotesi lontana, è fondamentale mantenere vivo il dibattito e l'attenzione sullo stato d'avanzamento dell'arte. Lo sviluppo tecnologico è rapidissimo e si insinua silenziosamente nelle maglie del processo. Il giudice potrebbe in futuro ritrovarsi relegato a un ruolo secondario, di mera legittimazione del risultato algoritmico, perdendo quel ruolo non solo di "arbitro" della controversia ma anche di esempio per la comunità sociale in cui opera.

---

[455] *Supra* nota 368, p. 65

[456] D. Polidoro, *Tecnologie informatiche e procedimento penale: la giustizia penale "messa alla prova" dall'intelligenza artificiale* in *Archivio Penale*, 2020, n. 3, p. 22

Su tali basi emergono le numerose criticità sollevate in merito alle decisioni che richiamano la discrezionalità del giudice: si pensi alle fattispecie rimesse al «prudente apprezzamento». La rigidità del modello matematico degli algoritmi è incompatibile con le operazioni compiute in sede giudiziale, tra cui la sussunzione del fatto nella norma, la selezione dei precedenti, la risoluzione delle antinomie o dei contrasti[457], il contemperamento dei valori della comunità, le ragioni delle parti e le circostanze accessorie più rilevanti[458]. I dispositivi di intelligenza artificiale non si rivelano certamente in grado di «compiere attività critiche, aventi ad oggetto, per esempio, il grado di coerenza delle dichiarazioni dei soggetti esaminati in sede processuale, le quali, pertanto, dovranno essere, in ogni caso, riservate al giudice-persona[459]».

Il necessario intervento dell'uomo si esprime anche nel momento della ricostruzione delle questioni di fatto, che devono essere rielaborate «in base alle narrazioni processuali dei soggetti coinvolti, cariche di percezioni, emozioni, punti di vista»[460]. Infatti, solo «l'ascolto empatico» che è proprio dell'umano è in grado di dare corpo e consistenza alle istanze di giustizia che devono essere riempite di senso, laddove invece una macchina «riduce la discrezionalità a un calcolo probabilistico[461]».

In conclusione, come sostenuto da Carnelutti[462], il diritto è «materia ribelle ai numeri[463]». La giustizia necessita di una dimensione dialettica, di un confronto in

---

[457] C. Casonato, *Intelligenza artificiale e giustizia: potenzialità e rischi in* DPCE *Online,* Vol. 44, n. 3, 2020

[458] T. Sourdin*, ult. op. cit.*, che prosegue sul punto affermando «Judicial commentary informs how society can operate and many judges also play a role in an educative sense, both informing litigants and lawyers about approaches to be taken and also contributing to civic education at a broader level. Proponents of the view that judges can be replaced by AI are arguably missing the point in relation to what judges contribute to society which extends beyond adjudication and includes important and often unexamined issues relating to compliance and acceptance of the rule of law.»

[459] D. Polidoro, *Tecnologie informatiche e procedimento penale: la giustizia penale "messa alla prova" dall'intelligenza artificiale* in *Archivio Penale*, 2020, n. 3, p.20. Sul punto *cfr.* Nieva-Fenoll, *Intelligenza artificiale e processo*, cit., 77

[460] L. Breggia, *op. cit*.

[461] C.V. Giabardo, *ult. op. cit.*

[462] Tra i più noti e autorevoli giuristi italiani.

[463] F. Carnelutti, *Matematica e diritto*, in Riv. Dir. Proc., 1951, p. 211-212

cui l'elemento umano non è sostituibile dalla macchina, di una valutazione, anche alla luce del proprio percorso umano e della propria sensibilità, delle sfumature varie e complesse della realtà, che un algoritmo non è certamente programmato per cogliere.

La preoccupazione è che venga gradualmente archiviato il diritto nel suo essere diritto *umano,* sostituito da valutazioni impersonali e frutto di calcoli statistici[464].

5.5 Il diritto a una sentenza individualizzata

Gli algoritmi sono elaboratori statistici che costruiscono modelli basati sul passato: tentano, cioè, di valutare la portata attuale o futura dei valori di una variabile partendo dall'analisi di esempi passati[465].

Il diritto a essere condannati sulla base di quanto commesso personalmente (e non sulla base di quanto hanno commesso altri in condizioni similari) è un diritto fondamentale per garantire l'equità di un processo.

È evidente che così il risultato, e dunque il giudizio, viene falsato dagli schemi comportamentali e dalle decisioni assunte in una determinata comunità di riferimento, contrariamente a quanto dettato dal principio di personalizzazione del trattamento sanzionatorio (desumibile anche dall'art. 27, commi 1 e 3, Cost.).[466]

L'algoritmo effettua una stima in relazione al gruppo economico-sociale cui è riconducibile l'imputato, basandosi su fattori statistici e caratteristiche immutabili e indipendenti dal soggetto (come il livello di istruzione e il quartiere di provenienza[467]).

Come osservato dall'ex Procuratore Generale degli Stati Uniti, Eric Holder, gli imputati sono ricondotti in gruppi e vengono assegnate condanne in base

[464] L.Avitabile, Introduzione al libro di B.Romano, *Algoritmi al potere: calcolo, giudizio, pensiero,* Giappichelli, Torino,2018
[465] CEPEJ, *Carta Etica*, par. 61
[466] V. Manes, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in *Discrimen*, 2020, pp. 13-14
[467] Former Attorney General Eric Holder's Speech at the 2016 Democratic National Convention

all'appartenenza a quei gruppi particolari, piuttosto che al crimine di cui sono stati condannati. Alcuni dei fattori utilizzati non sono nemmeno predittivi di un comportamento antisociale di per sè, ma sono condizioni sociali percepite con pregiudizio dalla società, ad esempio un basso livello d'istruzione o l'essere cresciuti con un genitore *single*.

Il giudizio algoritmico trascende dunque dalle singole azioni poste in essere dall'imputato e finisce per concentrarsi su fattori esterni che poco o nulla hanno a che fare con il rischio di recidiva. Si rischia di favorire una forma di determinismo penale per cui dal "diritto penale del fatto" – sancito anche dall'art. 25, comma 2, Cost. – si passa a un "diritto penale del profilo d'autore" nel quale «la pericolosità di un soggetto viene desunta esclusivamente dagli schemi comportamentali e dalle decisioni assunte in una determinata comunità del passato.[468]» In questo modo si favorisce la standardizzazione del singolo caso facendolo confluire nella statistica e le peculiarità del singolo accadimento che rischierebbero di sfumare nella nuvola delle probabilità.

5.5.1 Impatto dei dati statistici di massa nel giudizio: criticità

Gli strumenti di *Risk Assessment* confrontano le caratteristiche di un individuo con quelle di un gruppo di riferimento e lo score calcolato riflette il grado di somiglianza tra l'individuo e il gruppo: un *high-risk offender* condivide molte delle caratteristiche di soggetti che hanno recidivato in passato[469].

Sono state sollevate numerose perplessità dal punto di vista etico circa l'opportunità di assumere decisioni in merito alla libertà di un individuo sulla base di dati statistici di gruppo e condotte poste in essere da soggetti terzi.

Gli algoritmi, pur utilizzando le specifiche dell'individuo come *input*, li combinano con modelli contenenti dati di massa di altri gruppi sociali. Si svolge così un procedimento di sussunzione da dati collettivi per ragionare sul singolo caso

[468] V.Manes, *op. cit.*
[469] De Keijser, J.W., Roberts, J.V. and Ryberg, J. (eds.) (2019). *Predictive Sentencing: Normative and Empirical Perspectives,* Bloomsbury Publishing.

concreto, trasformando le correlazioni tra dati in fattori causali veri e propri. Si rischia così di considerare pericoloso un individuo «soltanto in virtù dell'appartenenza ad un gruppo: ciò produrrebbe l'indesiderato effetto di "contaminare" la valutazione discrezionale del magistrato senza fornire alcun elemento utile a desumere la capacità a delinquere del colpevole[470].»

Ulteriore aggravante è rappresentata dal fatto che tali dati sono basati su connotati immutabili dei singoli, contribuendo ad esacerbare le disparità socio-economiche e razziali anche all'interno del sistema carcerario.

Nel caso Loomis era stato sollevato come motivo di doglianza proprio il pregiudizio a ottenere una decisione conforme al principio del trattamento sanzionatorio individualizzato, cioè adeguato a quanto commesso dall'imputato. Nonostante la Corte del Wisconsin avesse disatteso i motivi del ricorrente, risulta evidente il pregiudizio del diritto a una sentenza individualizzata per via della previsione basata su dati generalizzanti e non attinenti all'individuo.

I *risk assessment tool* non stimano la probabilità specifica che quell'individuo possa recidivare, bensì producono una stima basata sulle similarità rispetto ad altri casi di soggetti che condividono quelle stesse caratteristiche. L'inserimento di variabili algoritmiche quali i precedenti penali ed il contesto familiare implica che la condotta passata di un certo gruppo «possa decidere il destino di una persona la quale, ovviamente, è un essere umano unico con un'origine sociale, un'istruzione, competenze specifiche, un grado di colpevolezza e motivazioni particolari per commettere un reato[471]».

In tal senso, la decisione automatizzata rischierebbe di essere tendenzialmente *conservatrice*, replicando gli schemi rilevanti del passato e non potendo immaginare scenari futuri.[472]

---

[470] L. D'agostino, ult. op. cit.
[471] CEPEJ, *Carta Etica*, Punto 134
[472] C. Casonato, *Intelligenza artificiale e giustizia: potenzialità e rischi* in *DPC Online,* [S.l.], v. 44, n. 3, oct. 2020, p. 3383

Così, gli imputati finiscono per esser giudicati non per quello che *hanno effettivamente commesso*, ma per quello che *potenzialmente potrebbero fare* in futuro[473]. Gli algoritmi prendono in considerazione dati in alcun modo collegati alla specifica criminale[474] e che non dovrebbero essere presi in considerazione in tali valutazioni (ad esempio, il livello di istruzione e il quartiere di provenienza) fondando la prognosi su fattori esterni e indipendenti dalla volontà del soggetto piuttosto che sulla vera responsabilità dello stesso.

Un ulteriore effetto collaterale determinato da tale meccanismo è la possibilità di assegnare un punteggio di rischio inferiore in caso di appartenenza a una classe sociale considerata "virtuosa": un livello di istruzione elevato, un quartiere residenziale con un basso tasso di criminalità, un'occupazione rispettabile aumentano sensibilmente le possibilità di ottenere un basso punteggio di rischio di recidiva. Ecco, dunque, che si predispone un terreno fertile per l'ingresso nel processo di dinamiche discriminatorie. Partendo da tali assunti, nel prossimo paragrafo si tenterà di delineare come tali strumenti facilitino l'ingresso a dinamiche discriminatorie in sede processuale.

5.6 Il divieto di discriminazione

Un altro principio frequentemente citato in relazione all'impiego degli algoritmi nel processo è il godimento dei diritti senza discriminazioni.

L'art.14 Cedu garantisce che il godimento dei diritti e delle libertà riconosciuti nella Convenzione sia assicurato «senza nessuna discriminazione, in particolare quelle fondate sul sesso, la razza, il colore della pelle, la lingua, la religione, le opinioni politiche o quelle di altro genere, l'origine nazionale o sociale, l'appartenenza a una minoranza nazionale, la ricchezza, la nascita od ogni altra condizione.»

[473] J. Nieva- Fenollm *op.cit.*, p 65
[474] *Ibidem*

Gli algoritmi possono diventare veicolo di fenomeni discriminatori: questi strumenti apparentemente neutri in realtà sottendono pregiudizi impliciti al loro interno. Gli strumenti di *machine learning* hanno il potenziale di rinforzare i pregiudizi esistenti poichè, a differenza degli umani, non sono in grado di rilevare consapevolmente e correggere i *bias* appresi e internalizzati[475], con il rischio di cristallizzare e moltiplicare esponenzialmente le discriminazioni. Si potrebbe fare l'esempio dell'incensurato residente in un quartiere dove risiedono a propria volta soggetti gravati da precedenti o caratterizzato da un elevato tasso di criminalità. È evidente che in assenza di opportuni correttivi d'istruzione dell'algoritmo, il soggetto potrebbe venire discriminato dall'assegnazione di un coefficiente di alto rischio di recidiva[476].

Come rilevato dal Consiglio d'Europa «*there is a danger that such systems perpetuate or exacerbate indirect discrimination through stereotyping. Indirect discrimination is only present where differential treatment cannot be justified*. »

Come già ampiamente dimostrato in riferimento agli strumenti di polizia predittiva, l'algoritmo può fornire risultati che rispecchiano "deviati": il massiccio invio di pattuglie in quartieri a maggioranza afroamericana o ispanica determina inevitabilmente un maggior numero di arresti nella zona. L'algoritmo, di conseguenza, effettuerà una associazione *razza-criminalità,* perpetrando la dinamica discriminatoria[477] .

Considerare fattori come l'educazione, il quartiere di provenienza e lo stato occupazionale come fattori predittivi determina uno squilibrio a danno dei gruppi sociali notoriamente più svantaggiati.

[475]Council of Europe*, Algorithms and Human rights: Study on the human rights dimensions of automated data processing techniques and possible regulatory implications,* Marzo 2018, consultabile su https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5
[476]A. Ziroldi, *Intelligenza artificiale e processo penale tra norme, prassi e prospettive* in *Questione di Giustizia*, 2019
[477] *Ibidem*

Al netto di tali considerazioni, acquista rilievo il concetto di *discriminazione indiretta*[478], fatto proprio anche dalla CEDU in numerose sentenze, secondo il quale «una differenza di trattamento può consistere nell'effetto sproporzionatamente pregiudizievole di una politica o di una misura generale che, se pur formulata in termini neutri, produce una discriminazione nei confronti di un determinato gruppo[479]». Il principio può essere esteso anche all'effetto discriminatorio del procedimento algoritmico: come dimostrato da ProPublica, a causa dei pregiudizi razziali contenuti nel COMPAS, era calcolata una probabilità di recidiva della popolazione nera *doppia* rispetto a quella della popolazione bianca nei due anni successivi alla condanna. I risultati dello studio verranno analizzati di seguito.

Alla luce degli studi condotti, risulta che determinati gruppi sociali sono colpiti in modo sproporzionatamente negativo rispetto ad altri gruppi in situazioni analoghe: pur essendo la procedura algoritmica apparentemente neutra, le previsioni di probabilità di recidiva più alte per gli afroamericani dimostrano effetti sproporzionati ai danni di tale categoria sociale. Per tutelare tali situazioni si potrebbe applicare per analogia il principio dell'inversione dell'onere della prova: se i dati dimostrano, per esempio, che le donne o le persone disabili sono particolarmente sfavorite, spetterà allo Stato fornire una spiegazione alternativa convincente di tali dati[480]. La CEDU lo ha chiarito nella causa *Hoogendijk c. Paesi Bassi*[481]: «[L]a Corte ritiene che, se un ricorrente è in grado di dimostrare, sulla base di statistiche ufficiali incontestate, l'esistenza di un indizio che una norma specifica — pur formulata in termini neutri — di fatto colpisca una percentuale di

[478] CEDS, Confederazione Generale Italiana del Lavoro (CGIL) c. Italia, ricorso n. 91/2013, 12 ottobre 2015, punto 237; Sul punto Agenzia dell'Unione europea per i diritti fondamentali e Consiglio d'Europa, 2019, *Manuale di diritto Europeo della non discriminazione*, 2018

[479] CEDU, Biao c. Danimarca [GC], n. 38590/10, 24 maggio 2016, punto 103; CEDU, D.H. e a. c. Repubblica ceca [GC], n. 57325/00, 13 novembre 2007, punto 184. La discriminazione indiretta può nascere qualora non si tenga conto di «tutte le differenze significative esistenti tra persone che si trovano in una situazione simile o di adottare le misure adeguate per garantire che i diritti e i vantaggi collettivi aperti a tutti siano realmente accessibili a tutti e da tutti.»

[480] Agenzia dell'Unione europea per i diritti fondamentali e Consiglio d'Europa, 2019, *Manuale di diritto Europeo della non discriminazione*, 2018

[481] Eur. Court of human rights, 1st Section, 6.1.2005, 58641/00, *Hoogendijk c. Paesi Bassi*

donne chiaramente più elevata rispetto agli uomini, spetta al governo convenuto dimostrare che ciò è il risultato di fattori oggettivi, non collegati a una discriminazione fondata sul sesso.»

5.6.1 L'illusione della neutralità: il pregiudizio implicito nell'algoritmo

Un rischio concreto legato all'uso degli algoritmi è rappresentato dalle proiezioni di pregiudizi razziali nel risultato fornito. Si distingue in tal senso tra pregiudizi "derivati" e pregiudizi "autonomi", a seconda che la distorsione si trovi:

- Nel codice sorgente, che potrebbe riflettere i preconcetti propri del programmatore o generati dai valori di riferimento dell'organizzazione in cui il programmatore opera[482] (si pensi all'inclusione o l'esclusione di caratteri che identificano o rinviano ad una categoria protetta, a rischio di discriminazione)[483].

- Nei *data* inseriti al suo interno per elaborare i modelli statistici, in quanto l'algoritmo utilizza dei *training data* per alimentarsi: tali dati possono facilmente riflettere *bias* di chi li ha selezionati o, più banalmente, contenere errori dovuti a generalizzazioni basate su dati incompleti, mezzi di raccolta non adeguati[484], dati parziali, incoerenti o non adeguatamente rappresentativi delle minoranze coinvolte[485]. Se il campione di dati raccolti risulta sensibilmente più ampio per un certo gruppo (ad esempio, gli afroamericani) e molto più esiguo per un altro, il gruppo sovra-rappresentato (o sotto-rappresentato) è sfavorito dalla falsa rappresentazione della realtà[486].

---

[482] P. Zuddas, *Intelligenza artificiale e discriminazion*i, in *Liber amicorum* per Pasquale Costanzo, 16 marzo 2020
[483] *Ibidem*
[484] Si pensi al caso in cui come modalità di acquisizione dei dati vengano utilizzati schedari di polizia in cui la percentuale di immigrati o afroamericani schedati risulta particolarmente elevata: il sistema imparerebbe che gli immigrati o gli afroamericani sono più inclini a commettere reati.
[485] S. Quintarelli, *ult. op*. cit. p.96
[486] P. Zuddas, *op. cit*., p.9

• In assenza di informazioni specifiche o correttivi in sede di programmazione, l'algoritmo potrebbe autonomamente individuare alcune caratteristiche che rinviano a categorie protette, associando ai loro detentori un trattamento deteriore[487].

I dati per l'apprendimento dell'algoritmo vengono etichettati manualmente dagli esseri umani: è evidente che, anche solo inconsciamente, vi è il rischio che in tale attività si possano riflettere giudizi di valori e preconcetti sociali.

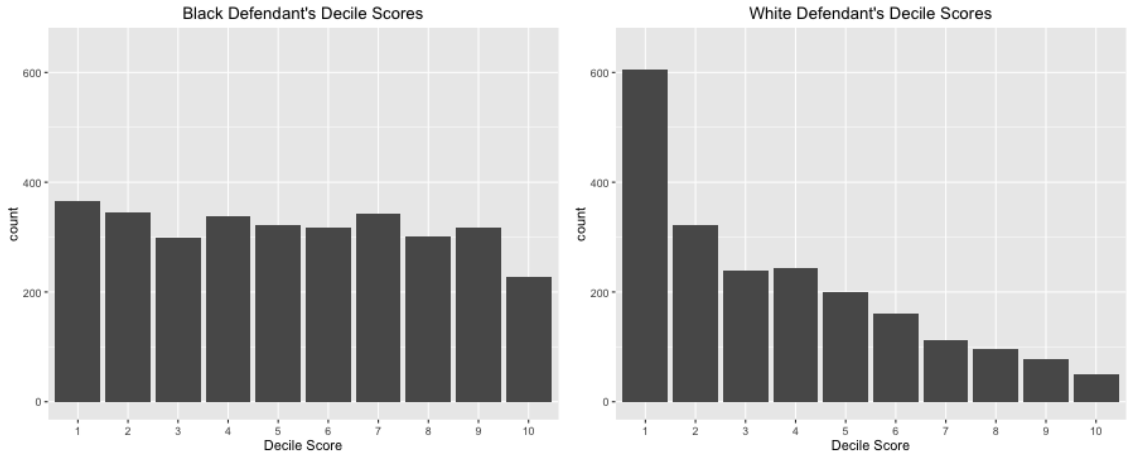
Individuare e correggere tali distorsioni non è possibile poiché i sistemi non sono trasparenti: emergeranno solo nel momento in cui gli individui lamentano la discriminazione subita: è impossibile effettuare un controllo *ex ante* a causa del vizio di opacità dell'algoritmo.

Nell'ambito della giustizia, i sistemi di valutazione del rischio si basano sui dati storici della giustizia penale, la qualità di tali proiezioni dipende dalla qualità dei dati del sistema di giustizia penale che sono stati utilizzati per svilupparli. Di conseguenza, i problemi persistenti con gli effetti dell'incarcerazione di massa, pratiche di polizia discriminatorie e altre discriminazioni nel sistema di giustizia penale conducono inevitabilmente le valutazioni del rischio a proiettare la stessa discriminazione e pregiudizio nel futuro.

A conferma di ciò, nel Maggio 2016 ProPublica ha pubblicato uno studio dal titolo *"Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks"*[488] in cui si dimostra come i risultati forniti dall'algoritmo fossero sistematicamente *racially biased*, determinando tassi di pericolo più elevati negli afro-americani rispetto ai bianchi, con una probabilità di più del doppio di essere individuati come ad "alto rischio" e il 45% in più di probabilità di commettere un qualsiasi tipo di crimine in futuro che sale al 77.3%

[487] *ibidem*
[488] J. Angwin, J. Larson, S. Mattu and L. Kirchner, *Machine Bias:There's software used across the country to predict future criminals. And it's biased against blacks.* in ProPublica, 2016 disponibile su https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

se si valuta il rischio di recidiva *violenta*. Gli imputati bianchi, al contrario, avevano maggiori probabilità (circa il 63%) di essere etichettati come a "basso rischio", ma poi hanno recidivato nei due anni successivi.



Gli istogrammi mostrano come le persone di colore siano sensibilmente più esposte al rischio di essere giudicate "ad alto rischio" di recidiva rispetto agli appartenenti al campione di popolazione bianco, a loro volta giudicati più spesso a "basso rischio" di circa il 63%. Nel campione c'erano 3.175 imputati neri e 2.103 imputati bianchi e del totale, 2.809 hanno recidivato entro due anni.

Il team di giornalisti di ProPublica ha comparato i punteggi di rischio[489] attribuiti dal COMPAS con gli effettivi casi di recidiva degli accusati nei successivi due anni (riscontrando un livello di precisione pari al 61% in caso di recidiva generale e del 21% in caso di recidiva violenta). Dividendo poi la popolazione in individui di colore e bianchi e confrontando i due campioni, il team rilevò che per gli individui di colore l'algoritmo prediceva un elevato numero di quelli che in statistica sono definiti *falsi positivi,* cioè soggetti classificati ad alto rischio che poi non commettevano alcun nuovo reato nei successivi due anni[490].

[489] ProPublica ha avuto accesso, in forza delle leggi sulla trasparenza, alle valutazioni di rischio fatte per oltre 7.000 persone arrestate nella contea di Broward in Florida tra il 2013 e il 2014

[490] A. Vespignani, *L'Algoritmo e l'oracolo: come la scienza predice il futuro e ci aiuta a cambiarlo*, Il Saggiatore, Milano, 2019, p. 106 e ss.

Per quanto attiene ai *falsi negativi,* come si osserva dalla tabella sottostante, sono soggetti per lo più appartenenti alla categoria dei bianchi. Nel caso dei soggetti a basso rischio, la probabilità che l'algoritmo ne etichetti erroneamente come pericoloso uno di colore è doppia rispetto a uno bianco. Viceversa, tra gli imputati che hanno recidivato, i detenuti bianchi che erano stati etichettati erroneamente come a basso rischio sono stati il doppio dei neri.[491]

	Bianchi	Afro-americani	
Etichettati come ad alto rischio, ma non hanno recidivato	23,5%	44,9%	
Etichettati come basso rischio ma hanno recidivato	47,7%	28%	

La compagnia Northepointe, Inc. ha replicato commentando che «*Northpointe does not agree that the results of your analysis, or the claims being made based upon that analysis, are correct or that they accurately reflect the outcomes from the application of the model*» e specificando che l'algoritmo produceva le previsioni con la stessa accuratezza per entrambi i campioni analizzati (*predictive parity*)[492]. Come ha osservato ProPublica, l'algoritmo dovrebbe, in ogni caso, identificare i soggetti ad alto rischio a prescindere dall'etnia di appartenenza.

Il pregiudizio riscontrato dall'analisi di ProPublica si è dimostrato correlato all'utilizzo di precedenti giudiziari (per lo più sfavorevoli per i condannati di colore), i quali avevano indotto il sistema a sovrastimare il rischio di recidiva per gli afroamericani[493].

[491] H.Fry, *op. cit*., p. 68 e ss.
[492] *Ibidem*
[493] P. Zuddas, *op.cit.*

Come dimostrato da numerosi studi, la neutralità degli algoritmi è un mito, in quanto i loro creatori, consciamente o meno, riversano in essi i loro sistemi di valori (tanto da essere stati definiti «un pregiudizio incorporato in un codice»[494]). Si consideri, inoltre, che anche laddove si fosse in grado di individuare la fonte del pregiudizio, a causa del complesso meccanismo di apprendimento dell'algoritmo risulterebbe pressoché impossibile riuscire a "rieducarlo" correttamente[495].

Dietro la loro facciata efficiente ed impersonale, i sistemi algoritmici rispecchiano le intenzioni di chi li progetta o li commissiona, generando un potere operativo e asimmetrico sulla vita di altre persone [496].

Al fine di scongiurare risultati discriminatori, è auspicabile estromettere informazioni quali il sesso, la razza e altri dati storicamente legati a dinamiche di segregazione sociale o elaborare sistemi di "filtraggio" selezionando accuratamente i dati da immettere nel sistema e prediligendo dati "neutrali", che non possano ricondurre il soggetto ad una determinata categoria.

5.6.2 Il rischio di cristallizzazione del pregiudizio: i feedback loops

È ormai indubbio che gli algoritmi siano "casse di risonanza" di pregiudizi sociali che si riflettono nell'*outcome* prodotto: le decisioni fondate su statistiche storiche perpetrano e contribuiscono a cristallizzare i pregiudizi in esse incorporate.

[494] C. O'Neil, *op. cit.*
[495] A. Jean, *Nel paese degli algoritmi,* Neri Pozza editore*,* Vicenza*,* 2021, p. 112
[496] E. Sadin, *ult. op. cit.;* Tale rilievo è condiviso dal criminologo Aleš Zavrsnik che sottolinea che le fasi di costruzione e interpretazione degli algoritmi sono «prodotte da uomini per uomini e, comunque questi ultimi siano concepiti, non possono sfuggire agli errori, ai pregiudizi, agli interessi umani e alla rappresentazione umana del mondo. »

Gli algoritmi rischiano così di innescare dei "circoli viziosi"[497] (c.d. *feedback loops*[498]), proprio perché vengono alimentati con i dati che producono loro stessi. Nel caso dei *software* di polizia predittiva, ad esempio, inviare pattuglie sul posto implica la possibilità di raccogliere nuovi dati, di conseguenza, ciò incrementerà il numero dei crimini registrati e giustificherà nuovi interventi di polizia sul posto[499]. Il problema sotteso è ben spiegato: le zone *ad alto rischio* potrebbero essere considerate tali solo perché la polizia ha più dati a disposizione riguardo quella specifica area o comunità rispetto ad altre.[500]

Come più volte evidenziato, l'algoritmo contribuisce ad esacerbare dinamiche discriminatorie, ad esempio determinando un sovra-pattugliamento delle aree a maggioranza afroamericana o ispanica. L'algoritmo, così, arriverà a ritenere corretta l'associazione *razza-criminalità*, inviando così la polizia solo nei quartieri a maggioranza di persone di colore[501] come in una profezia auto-avverante. A titolo di esempio, se un dipartimento utilizza un algoritmo in cui sono stati introdotti

---

[497] Uno studio condotto su PredPol ha evidenziato che inviare pattuglie di polizia nei quartieri indicati dall'algoritmo incrementa i crimini registrati e del 20%. Registrando tale incremento nell'algoritmo, lo stesso «became orders of magnitude more confident that its predictions were correct. '[This] creates a feedback loop, [in which] the algorithm becomes more certain about these places that are over-policed" ». In argomento Renata M. O'Donnell, *'Challenging Racist Predictive Policing Algorithms under the Equal Protection Clause'* (2019) ,94(3) New York University Law Review 544, p. 563

[498] Per avere un'idea del fenomeno, basta pensare alle società che ricorrono a sistemi automatici per la concessione del credito sulla base di un certo punteggio di affidabilità: se il richiedente proviene da un quartiere economicamente svantaggiato, con alti tassi di soggetti inadempienti, saranno richiesti tassi di interesse più elevati. Ciò può portare a un ciclo di feedback che rafforza le pratiche di prestito discriminatorie esistenti: se i richiedenti hanno difficoltà a pagare queste commissioni più elevate, questo indica al sistema che erano effettivamente a rischio più elevato, il che si tradurrà in punteggi inferiori per altri richiedenti simili in futuro. Sul punto C. O' Neil, op. cit. p. 156 e ss.

[499] C. O 'Neil, op.cit., p.128; in argomento A. Mittone, *Giustizia digitale* in Doppiozzero, 9 marzo 2020 consultabile su https://www.doppiozero.com/materiali/giustizia-digitale

[500] Ferguson, Andrew Guthrie. "*Policing Predictive Policing*." Washington University Law Review, vol. 94, no. 5, 2017, p. 1149

[501] «Under this analysis, predictive policing algorithms will learn less about crime in predominantly white areas and will report that there is less of a risk of future crime in those areas, while learning more about predominantly Black neighborhoods and indicating that more police personnel should be sent to those areas». In argomento *supra* nota 229.

biased data, le previsioni di rischio possono condurre a maggiori arresti nei confronti di soggetti appartenenti ad una minoranza, e di conseguenza si amplifica la probabilità che quella minoranza sia considerata pericolosa.

Tale rischio viene colto ed evidenziato anche all'interno della Carta Etica: «i quartieri considerati a rischio attirano maggiormente l'attenzione della polizia, la quale scopre conseguentemente un maggior numero di reati, con il risultato di un'eccessiva sorveglianza da parte della polizia delle comunità residenti in tali luoghi.»[502]

Un altro studio, condotto dallo *Human Rights Data Analysis Group* su PredPol, ha evidenziato come l'algoritmo abbia utilizzato i dati della polizia per generare previsioni di crimini di droga ad Oakland, e di come questo abbia raccomandato di indirizzare il doppio delle risorse di polizia nelle aree "nere" rispetto alle aree "bianche", nonostante i reati di stupefacenti fossero ragionevolmente distribuiti equamente tra i due gruppi. Si innescano, in tal modo, dei circuiti grazie ai quali la polizia viene ripetutamente rimandata nello stesso quartiere in un modo che rafforza ed esacerba le distorsioni iniziali nei dati di addestramento (dati di *input*).

Le disparità di previsione osservate nello studio pubblicato da ProPublica non sono frutto di una formulazione statistica fallace, ma il «*mathematical result of the divergent rates of arrest between the black and white defendants in the underlying dataset (...) so long as the algorithm is also striving to have equal predictive accuracy [calibration] for each racial group*.[503]»

Risulta chiaro che, soprattutto per gli algoritmi utilizzati in fase di modulazione della sentenza, vi sia un elevato rischio di discriminazione per gli appartenenti ad una minoranza etnica, sensibilmente più esposti ad una maggior probabilità di essere etichettati come ad "alto rischio" (anche se non hanno recidivato nei due anni successivi). In questo modo, si alimenta il ciclo di incarcerazione legata alle condizioni di svantaggio socio-economico, cristallizzando i pregiudizi nei confronti

[502] CEPEJ, Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi, par. 121

[503] S. Mayson, *Bias In, Bias Out*, 128 YALE L. J. 2218, 2234 (2019).

delle minoranze e le disuguaglianze sociali alla base di numerosi comportamenti criminali che verranno perpetrati dall'algoritmo.

5.7 La presunzione di innocenza

Specifica garanzia processuale prevista dall'Art. 6 para. 2 è la presunzione di innocenza, secondo cui «ogni persona accusata di un reato è presunta innocente fino a quando la sua colpevolezza non sia stata legalmente accertata.»

La norma produce effetti su più livelli: come *regola di trattamento,* in quanto l'imputato deve essere trattato formalmente e sostanzialmente come tale, fino a che la sua colpevolezza non venga legalmente accertata con sentenza definitiva; come *regola probatoria*, per cui l'onere della prova è distribuito tra le parti, ovvero incombe sull'accusa l'onere di dimostrare la colpevolezza dell'imputato; infine, come *regola di giudizio*, in quanto se la colpevolezza non è dimostrata al di là di ogni ragionevole dubbio, l'imputato deve essere assolto[504].

Il principio di presunzione d'innocenza impone che i membri di un tribunale non partano con la convinzione precostituita che l'imputato abbia commesso il fatto addebitatogli. In secondo luogo, la norma prescrive che l'onere della prova sia posto a carico dell'accusa[505]: la presunzione d'innocenza potrà dirsi violata nel caso in cui l'onere della prova sia invertito, facendolo gravare indebitamente sulla difesa e non sull'accusa[506]. Ammettendo gli algoritmi nel processo penale, invero, si rischia di invertire l'onere della prova: come ampiamente dimostrato[507], l'utilizzo dei *risk assessment tool* possono indebitamente influenzare la percezione dell'innocenza o della colpevolezza dell'imputato[508].

Certamente, quello enunciato non è un diritto assoluto, in quanto le presunzioni di colpevolezza operano in qualsiasi sistema penale, a patto che però gli argomenti siano ragionevolmente *confutabili*. Come dimostrato, il vizio di opacità algoritmico

[504] V. Zagrebelsky, *op.cit.*
[505] Consiglio d'Europa / Corte Europea dei Diritti dell'Uomo, *Guida all'art 6*, 2014
[506] Eur. Court of human rights, 3rd Section , 20.3.2001, 33501/96, *Telfner v. Austria*
[507] *Supra* para. 6, cap. V
[508] J.Nieva- Fenoll, *ult. op. ci*t., p. 142

non rende scrutinabile né il procedimento logico seguito dal software, né la qualità dei dati inseriti ed il peso specifico attribuito ad ognuno di essi, rendendo l'onere sproporzionato ai danni della difesa. Inoltre, i *data* inseriti non sono neutrali, ma riflettono a vari livelli i pregiudizi del programmatore o dei soggetti deputati a selezionare i dati per elaborare i modelli statistici[509].

Tramite la raccolta indiscriminata di dati del soggetto, si elabora un profilo dello stesso, riconducendolo ad una determinata categoria sociale. Se l'imputato è, a titolo di esempio, un uomo afro-americano che proviene da un quartiere residenziale con un alto tasso di criminalità e con un basso livello di istruzione, è probabile che l'algoritmo di valutazione del rischio assegnerà un alto punteggio di recidiva, incoraggiando il giudice ad escludere misure alternative alla detenzione. Si creano così profili di "persone sospette" che di per sé non è una novità. Il problema risiede nel fatto che tali profili si creano sulla base di una selezione dei dati "viziata" e di parte[510].

Non è fattibile individuare l'autore di un crimine solo sulla base di una serie di caratteristiche esterne: la condanna deve essere basata esclusivamente sui fatti commessi, in quanto «non si può imporre una pena se l'unico strumento che abbiamo è il suo potenziale profilo psicologico, per altro non completamente affidabile[511].»

L'elaborazione di tali profili, se usati per applicare una misura cautelare o modulare una sentenza, finiscono con il violare la presunzione di innocenza: l'accusato partirà già in una condizione di svantaggio poiché il giudice sarà condizionato da tali fattori esterni e non da elementi probatori concreti.

[509] Consiglio d'Europa, Corte Europea dei Diritti dell'Uomo, *Guida all'art 6*, 2014
[510] J. Nieva- Fenoll, *op. cit*.. p. 142 e ss.
[511] *Ibidem*

CONCLUSIONI

L'incertezza che caratterizza la nostra epoca ha spinto l'uomo a ricorrere sempre più frequentemente a strumenti che possano, in qualche misura, ridurla. Gli algoritmi si pongono in tal senso come un *pharmakòn* capace di contrastare la fallibilità delle valutazioni umane. Nella società della "prestazione", in cui l'imperativo è quello dell'efficienza ad ogni costo, gli algoritmi trovano un terreno fertile di crescita ed espansione in ogni campo, sino al punto che è ormai impossibile immaginare un ambito del quotidiano in cui non trovino applicazione.

Nel corso della trattazione si è cercato di dimostrare il carattere ambivalente di tali strumenti: sarebbe riduttivo e fin troppo semplicistico prendere una posizione netta di contrapposizione rispetto ad essi, anche perché, come dimostrato, i benefici apportati sono molteplici.

Ben si comprende, dunque, come l'uso degli algoritmi nell'ambito della giustizia non vada demonizzato a prescindere, ma necessiti di una regolamentazione puntuale da parte dell'ordinamento. La sinergia tra uomo e macchina affascina da secoli l'essere umano e può costituire un valido supporto alle attività decisionali, ma occorre fissare dei limiti e delle linee-guida uniformi a livello internazionale, per evitare che i grandi benefici lascino spazio alle gravi violazioni dei diritti fondamentali.

Si è dimostrato come l'impiego dei *risk assessment tool* in fase di *presentencing* e *sentencing* nell'ordinamento statunitense, si risolva il più delle volte in una surrettizia elusione delle garanzie dell'equo processo. La *deregulation*, i risultati non scrutinabili, gli implicit *bias* contenuti nei *software* algoritmici e il rischio di delega della decisione all'algoritmo (ritenuto neutrale e oggettivo in virtù *dell'automation bias*) rischiano di minare in radice le garanzie summenzionate in assenza di un controllo esterno.

Siamo ormai entrati in una nuova era di *tecnoumanesimo* ed è auspicabile ripensare i rapporti tra uomo e macchina, nei quali le prestazioni cognitive del primo vengano potenziate dall' intelligenza artificiale ed eventualmente, verificare le

incongruenze ed i contrasti del processo decisionale umano, senza però arrivare a sostituirlo[512].

I limiti regolamentari e normativi sono dunque funzionali e necessari a realizzare questa interazione feconda e rispettosa dei diritti, soprattutto in fase di giudizio: l'uomo si affida a scelte etiche, la macchina a scelte efficienti e coerenti, che tuttavia non sempre assicurano l'equità, come più volte dimostrato in corso di trattazione[513]. La macchina non può cogliere le sfumature complesse di ogni situazione, non può effettuare un bilanciamento di giudizi e una gradazione degli interessi in gioco: l'attività decisionale umana è infungibile in tal senso. Il rischio è che in questo modo il giudice *decide* senza propriamente *giudicare*.

Al netto di tali considerazioni, è solo tramite un approccio normativo che si potrà scongiurare il rischio di trasformare un enorme potenziale di crescita in un'arma lesiva delle garanzie e della dialettica processuale.[514] Come scrive lo storico Yuval N. Harari nel suo saggio di attualità *"21 Lessons for the 21st Century" «*Gli esseri umani sono sempre stati di gran lunga più bravi a inventare strumenti che a usarli con saggezza*.»*

[512] A. *Punzi, Judge in the Machine: e se fossero le macchine a restituirci l'umanità del giudicare?* In A. Carleo (a cura di), *op. cit.*

[513] Come anche osservato dalla Vicepresidente della LUISS Guido Carli Paola Severino in una recente intervista rilasciata al Messaggero dal titolo «*Le macchine sono un dono ma servono regole globali*»

[514] Non a caso di recente sono stati elaborati due documenti in tal senso: il primo proviene dalla National Security Commission on Artificial Intelligence, accompagnato da una piena e rara intesa tra Congresso e Presidenza americana. Il secondo, proviene dalla Committee ad hoc on Artificial Intelligence istituito dal Consiglio d'Europa. Entrambe richiamano la necessità di un approccio multidisciplinare e regolamentare al fenomeno

BIBLIOGRAFIA

ALEVEN, V., ASHLEY K. *Evaluating a learning environment for case-based argumentation skills* in *ICAIL '97: Proceedings of the 6th international conference on Artificial intelligence and law ,*1997

ANDREJEVIC M., *Data collection without limits* in *Big Data, Crime and Social Control (Aleš Završnik ed., 2018)*

ARCEIRI S., *Bias cognitivi e decisione del giudice: un'indagine sperimentale* in *Diritto Penale e Uomo,* 2/2019.

BANERJEE R., *Estonia develops "robot judge"* in *New statesman,* Vol.148(5474), 2019

BRANTHINGHAM P.J., *The Logic of Data Bias and its Impact on Place-Based Predictive Policing* in *Ohio State Journal of Criminal Law*, 2018

BURCHARD C*., L'intelligenza artificiale come fine del diritto penale? Sulla trasformazione algoritmi della società* in *Rivista italiana di diritto e procedura penale*, Vol. 62, N° 4, 2019

BURREL J., *How the machine 'thinks': Understanding opacity in machine learning algorithms* in *Big Data & Society*. June 2016

CAPLAN J.M., KENNEDY L.W., BARNUM J.D., PIZA E.L, *Crime in Context: Utilizing Risk Terrain Modeling and Conjunctive Analysis to Explore the Dynamics of Criminogenic Behavior Setting* in *Journal of Contemporary Criminal Justice,* 33(2), 2017

CARCATERRA A., *Machinae autonome e decisione robotica* in in CARLEO A. (a cura di) *Decisione Robotica, Il Mulino,* Bologna, 2019

CARNELUTTI F.,*Matematica e diritto*, in Riv. Dir. Proc., 1951

CASTELLI C., PIANA D., *Giusto processo e intelligenza artificiale*, Maggioli Editore, Santarcangelo di Romagna, 2019.

CASONATO C., *Intelligenza artificiale e giustizia: potenzialità e rischi in* DPCE *Online ,* Vol. 44, n. 3, 2020

CESARI C., *Editoriale: L'impatto delle nuove tecnologie sulla giustizia penale – un orizzonte denso di incognite* in *Revista brasileira de direito processual penale, Porto Alegre,* Vol. 4, n. 3

CHIARELLI M., *Intelligenza artificiale e regolazione: problematiche e prospettive* in *Amministrazione in Cammino,* 2020

CHANENSON S. L., HYATT J. M., *The Use of Risk Assessment at Sentencing: Implications for Research and Policy,* in *Villanova University Charles Widger School of Law Public Law and Legal Theory*, 2016

CITRON D., *Technological Due Process* in *Washington University Law Review,* Vol.85, 2008

D' AGOSTINO L.,*Gli algoritmi predittivi per la commisurazione della pena* in *Diritto Penale Contemporaneo,* n. 2/2019

DE KERKHOVE D., La decisione datacratica, in CARLEO A. (a cura di) *Decisione Robotica*, Il Mulino, Bologna, 2019

DE MIGUEL BERIAIN I., *Does the use of risk assessments in sentences respect the right to due process? A critical analysis of the Wisconsin v. Loomis ruling* in *Law, Probability and Risk* (2018)

DE RENZIS L., *Primi passi nel mondo della giustizia «High Tech»: La decisione in un corpo a corpo virtuale fra tecnologia e umanità* in CARLEO A. (a cura di) *Decisione Robotica,* Il Mulino, Bologna, 2019

DE KEIJSER J., ROBERTS J.W., RYBERG J. *Predictive Sentencing: Normative and Empirical Perspectives,* Bloomsbury Publishing, 2019

DONATI F., *Intelligenza artificiale e giustizia* in *Rivista Associazione italiana dei Costituzionalisti* n. 1/2020,

DONOHUE M. E., *A replacement for justitia's scale? Machine learning in sentencing* in *Harvard Journal of Law & Technology*, Volume 32, Number 2, 2019

DI PRISCO A., *Elementi di criticità sulla perizia psicologica nel processo penale*, in *Ius in Itinere,* 2018

DIAKOPOULOS N., *Algorithmic Accountability Reporting: On the Investigation of Black Boxes* in *Tow Center for Digital Journalism*, Columbia University, 2014

EK. C.H., MALIK I., *Neural Translation of Musical Style* in *arXiv*, Agosto, 2017

EAGLIN J., *The Perils of 'Old' and 'New' in Sentencing Reform* , NYU Annual Survey of American Law, Forthcoming, 2020

FRY H., *Hello Word,* Bollati Boringhieri, Torino, 2019.

FERGUSON G., *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, NYU Press, 2017

FREEMAN K., *Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State v. Loomis* in *North Carolina Journal of Law & Technology*, Vol. 18, 2016

GALETTA D.U., CORVALAN J. G., *Intelligenza Artificiale per una Pubblica Amministrazione 4.0? Potenzialità, rischi e sfide della rivoluzione tecnologica in atto,* in *Federalismi*, n.3/ 2019

GARDNER H., *Frames of mind: The theory of multiple intelligences*, Basic Books, New York, 2011

GERARDAS J., T*he fundamental rights challenges of algorithms* in *Netherlands Quarterly of Human Rights,* Vol. 37(3), 2019

GIALBARDO C.V., *Il giudice e l'algoritmo (in difesa dell'umanità del giudicare)*, in *Giustizia Insieme*, 2020

GIALUZ M., *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assesssment tools tra Stati Uniti e Europa* in *Dir. pen. cont.,* 29 maggio 2019

GRAMMELGARD M., KOIVISTO A., ERONEN M., KALTIALA-HEINO R., *The predictive validity of the Structured Assessment of Violence Risk in Youth (SAVRY) among institutionalised adolescents* in *The Journal of Forensic Psychiatry & Psychology*, Vol. 19, 2008

GRISSO T., VINCENT G., SEAGRAVE D., *Mental Health screening and assessment in juvenile justice* in C. L. Kessler & L. J. Kraus (Eds.), *The mental health needs of young offenders: Forging paths toward reintegration and rehabilitation* , Cambridge University Press

GUO P., KEHL D., KESSLER S., *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing. Responsive Communities Initiative, Berkman Klein Center for Internet & Society* in Harvard Law School, 2017

GUTHRIE, C.; RACHLINSKI J., WISTRICH, A. J., *Blinking on the Bench: How Judges Decide Cases* (2007), *Cornell Law Faculty Publications*, Paper 917

HARARI N. Y., 21 *Lessons for the 21th Century*, Jonathan Cape, Londra, 2018

HAMILTON M., *Predictive Policing through risk assessment* in MCDANIEL J., PEASE K., *Predictive Policing and Artificial Intelligence,* Taylor and Francis, Milton Park, 2021

PEASE K., *Predictive Policing and Artificial Intelligence,* Taylor and Francis, Milton Park, 2021

HARCOURT B., *Against prediction: Profiling, policing and punishing in actuarial age*, University of Chicago Press, Chicago,2007

HUQ A. Z., *Racial Equity in Algorithmic Criminal Justice*, in *Duke Law Journal, V.68. n° 6, 2019*

JEAN A., *Nel paese degli algoritmi,* Neri Pozza editore*,* Vicenza*,* 2021

LUM K., ISAAC W.*, To Predict and Serve?* In *Significance* Vol. 13, no. 5, 2016

MALDONATO L., *Algoritmi predittivi e discrezionalità del giudice: una nuova sfida per la giustizia penale* in *Diritto penale contemporaneo*, Vol. 2/2019

MARINUCCI G., DOLCINI E. GATTA G.L., *Manuale di diritto penale-Parte Generale,* Giuffrè , Milano , 2019

MARMO R., *Algoritmi per l'intelligenza artificiale. Progettazione dell'algoritmo – Dati e Machine Learning - Neural Network - Deep Learning*, Hoepli, Milano, 2020

MAYSON S., *Bias In, Bias Out*, in YALE Law Journal, Vol. 128, n°8, 2019

MCGREGOR L., MURRAY D., *International human rights law as a framework for algorithmic accountability* in *International and Comparative Law Quarterly,*Vol. 68, 2019

MEYERS J.R., SCHMIDT F., *Predictive Validity of the Structured Assessment for Violence Risk in Youth (SAVRY) With Juvenile Offenders* in *Criminal Justice and Behavior* 35, no. 3 (March 2008)

NIEVA- FENOLL J., *Intelligenza artificiale e processo,* Giappichelli Editore, *Torino, 2019.*

NISSAN E., *Digital technologies and artificial intelligence's present and foreseeable impact on lawyering, judging, policing and law enforcement*, Springer-Verlag, London , 2015

O' NEIL C. *Armi di distruzione matematica. Come i Big Data aumentano la disuguaglianza e minacciano la democrazia,* Bompiani, Firenze, 2017.

O'DONNELL R. M, *Challenging Racist Predictive Policing Algorithms under the Equal Protection Clause* in *New York University Law Review* 544 (June 2019)

OSWALD M., GRACE J., URWIN S. , BARNES G.C., *Algorithmic Risk Assessment Policing Models: Lessons from the Durham HART Model and 'Experimental' Proportionality*, in *Information & Communications Technology Law*, n. 27, 2018

PALAZZANI L., *Tecnologie dell'informazione e intelligenza artificiale: sfide etiche al diritto,* Edizioni Studium S.r.l., 2020

PARODI C., SELLAROLI V., *Sistema penale e Intelligenza Artificiale: molte speranze e qualche equivoco* in *Diritto penale contemporaneo*, n. 6/2019

PASQUALE F.. *Paradoxes of Privacy in an Era of Asymmetrical Social Control* in *Big Data, Crime and Social Control* (Aleš Završnik ed., 2018)

PASQUALE F., *The black box society: The secret algorithms that control money and information,* Harvard University Press, 2015

PERES E., *Che cosa sono gli algoritmi*, Salani Editore, Milano, 2020

PERRY W. L., MCINNIS B., PRICE C. C., SMITH S. C., HOLLYWOOD J. S.. *Predictive Policing:* The Role of Crime Forecasting in Law Enforcement Operations. RAND Corporation, 2013.

PESCE G., *Il Consiglio di Stato ed il vizio della opacità dell'algoritmo tra diritto interno e diritto sovranazionale,* in *Giustizia Amministrativa*, 2020

PIERGALLINI C., *Intelligenza Artificiale: da mezzo a autore del reato?* In *Rivista Italiana di Diritto e Procedura Penale*, n. 4 del 2020

POLIDORO D., *Tecnologie informatiche e procedimento penale: la giustizia penale "messa alla prova" dall'intelligenza artificial*e in *Archivio Penale*, 2020, n.3

PUSTORINO P., *Lezioni di tutela internazionale dei diritti umani*, Cacucci Editore, Bari,2019

QUINTARELLI, S., *Intelligenza Artificiale: cos'è davvero, come funziona, che effetti avrà,* Bollati Boringhieri, Torino, 2020

QUATTROCOLO S.*, Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale "predittiva"* in *Cassazione penale* n .4/2019

QUATTROCOLO S., ANGLANO C., CANONICO M., GUAZZIONE M., *Technical Solutions for Legal Challenges: Equality of Arms in Criminal Proceedings* in *Global Jurist 20,* 2020

QUATTROCOLO S., *Processo penale e rivoluzione digitale: da ossimoro a endiadi?* In *Medialaws,* 3/2020

QUATTROCOLO S., *Equità del processo penale e automated evidence alla luce della Convenzione europea dei diritti dell'uomo* in *Revista Ítalo-Española de Derecho Procesal,* Vol. 1 | 2019

RUSSEL J., NORVIG P., A*rtificial Intelligence. A Modern Approach*, Third Edition, Prentice Hall, 2010

ROMANO B., *Algoritmi al potere: calcolo, giudizio, pensiero,* Giappichelli, Torino, 2018

SADIN E., *Critica della ragione artificiale: una difesa dell'umanità,* Luiss University Press, Roma, 2019

SEARLE J., *Minds, Brains and Programs*, in *Behavioral and Brain Sciences*, 1980

SERRA C,*Il diritto di contestazione delle decisioni automatizzate nel GDPR* in *Anuario de la Facultad de derecho de la Universidad de Alcalà*, Vol. XII, 2019

SPIELKAMP M, *Automating Society: Taking Stock of Automated Decision-Making* in the *EU BertelsmannStiftung Studies,* 2019

SPIEGELHALTHER D.J.*, The Future lies in Uncertainty*, in *Science*, 2014, vol. 435

SCHUILENBURG, M.., PEETERS R., *The Algorithmic Society: Technology, Power, and Knowledge*, Routledge, 2020

STEVENSON M., *Assessing Risk Assessment in Action in Minnesota Law Review.* V. 103, 2018

TALIA D., *La società calcolabile e i Big Data: algoritmi e persone nel mondo digitale*, Rubbettino Editore, Soveria Mannelli, 2018

TILLER L., *A Minority Report: The Unregulated Business of Automating the Criminal Justice System in The Business* in *Entrepreneurship & Tax Law Review's B.E.T.R. White Paper,* Marzo 2019

VARZI A. C., *L'intelligenza e l'artificiale*, in *KOS. Rivista di Scienza e Etica*, 1991

VESPIGNANI A., *L'algoritmo e l'oracolo: come la scienza predice il futuro e ci aiuta a cambiarlo*, Il Saggiatore, Milano, 2019

ZAGREBELSKY V., *Manuale dei diritti fondamentali in Europa*, Il Mulino, Bologna, 2019

ZIROLDI A,, *Intelligenza artificiale e processo penale tra norme, prassi e prospettive*, in *Quest. Giust.,* 18 ottobre 2019

ZUDDAS P., *Brevi note sulla trasparenza algoritmica* in *Amministrazione in Cammino*, 2020

INDICE DELLA GIURISPRUDENZA INTERNA

INDICE DELLA GIURISPRUDENZA SOVRANAZIONALE