

IANUS

Diritto e Finanza



Quaderni

<https://www.rivistaianus.it>



ISSN: 1974-9805

Quaderni - 2023

INTELLIGENZA ARTIFICIALE
E DECISIONE PENALE:
QUALI GLI SCENARI POSSIBILI?

Marzia Gaia Marzano

INTELLIGENZA ARTIFICIALE E DECISIONE PENALE: QUALI GLI SCENARI POSSIBILI?

Marzia Gaia Marzano

*Dottoranda di ricerca in Scienze giuridiche
Università degli Studi di Foggia*

Dopo una sommaria esposizione del funzionamento dei sistemi di Intelligenza artificiale (IA), il testo si concentra sui vantaggi e sui rischi legati all'utilizzo dell'IA nella decisione penale, tenendo conto delle politiche adottate in ambito comunitario e degli eventuali ostacoli al loro impiego che derivano dal diritto interno. Gli attuali limiti dei sistemi di IA, unitamente all'impossibilità di scongiurare il pericolo che si realizzi un conformismo generalizzato all'output emesso dall' algoritmo, inducono ad escludere un'interazione tra sistemi intelligenti e decisione penale con riferimento alla fase dell'accertamento dei fatti e dell'an della responsabilità. Il lavoro, tuttavia, giunge a conclusioni parzialmente diverse con riferimento alle valutazioni prognostiche che spesso il giudice penale è chiamato a svolgere aprendo all'apporto dell'IA nel caso in cui il progresso tecnologico riesca a superare gli ostacoli ad oggi presenti.

The Author focuses on benefits and risks associated with the use of artificial intelligence (AI) in criminal decisions, after explaining the function of AI systems, taking into account the EU policies and possible obstacles which can arise from the domestic legislation. The current limits of AI systems and the impossibility to avert the danger of a general conformation to algorithm's output exclude the interaction between intelligent systems and final criminal decision, in relation to the establishment of facts and accountability. However, this paper comes to partially different conclusions with regard to the assessment, which the criminal judge is called upon to perform by opening to the AI contribution in the case in which the technological progress will fix the actual issues.

Sommario:

1. Intelligenza artificiale e diritto penale: delimitazione del campo d'indagine
2. Tecnologia e giustizia penale: dalle banche dati ai software intelligenti. Cosa si intende per giustizia predittiva?
3. I vantaggi del ricorso alla giustizia predittiva...
4. ...ed i rischi legati al suo utilizzo
5. Le politiche adottate in ambito comunitario: il ridimensionamento dei rischi legati all'IA in fase decisoria
6. Osservazioni conclusive

1. Intelligenza artificiale e diritto penale: delimitazione del campo d'indagine

L'attuale epoca si sta caratterizzando per una pluralità di mutamenti in parte causati dalla rivoluzione digitale che incide in pressoché ogni attività.

Nonostante alcune criticità, può dirsi che il sistema penale sia stato fino a questo momento in grado di reggere l'urto dato dal sempre maggiore utilizzo di strumenti tecnologici.

Da un po' di anni a questa parte, però, l'ultima frontiera della tecnologia, rappresentata dalla intelligenza artificiale (IA), preoccupa oltremodo i giuristi, i quali temono le ripercussioni che l'impatto della digitalizzazione può avere sulla tenuta del sistema penale, sia dal punto di vista sostanziale sia da quello processuale.

Più nello specifico, il dibattito penalistico si sta concentrando su alcuni filoni di indagine che possono essere così sintetizzati¹.

Da un lato, quello che attiene ai problemi che la diffusione della IA potrà produrre in materia di imputazione della responsabilità quando provochi o contribuisca a causare un reato; dall'altro quello che riguarda l'apporto che tali tecnologie possono fornire ai sistemi di giustizia penale.

È bene, tuttavia, precisare che i problemi posti dall'IA in tale ultimo settore sono molti e diversi tra loro, sicché si ritiene opportuno distinguere a seconda che l'attenzione si concentri sul contributo che i sistemi intelligenti possono fornire in materia di prevenzione di reati² ovvero di accertamento della responsabilità penale.

Con riferimento a questo tema, suole ulteriormente distinguersi a seconda che ci si riferisca ai problemi posti dalle c.d. *digital evidence* basate sull'IA ovvero a quelli sollevati dall'impiego di *software* intelligenti nell'ambito dell'attività decisoria (in senso lato).

Il primo dei problemi menzionati impone di verificare se e a che titolo le risultanze probatorie ottenute attraverso tali strumenti possano essere acquisite nel processo penale³; il secondo, invece, è destinato a coinvolgere

¹ L'elencazione riportata riproduce quella elaborata da UBERTIS, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, in DI ARGENTINE (a cura di), *Giurisdizione penale, intelligenza artificiale ed etica del giudizio*, Milano, 2021, p. 17; ID, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, in *Dir. Pen. cont.* (Riv. Trim.), 4, 2020, p. 80 il quale fa riferimento anche allo scenario in cui la macchina sia vittima di reato.

² Il riferimento è ai problemi posti dai sistemi di IA chiamati a coadiuvare l'autorità di pubblica sicurezza nell'attività di prevenzione della commissione di fatti penalmente rilevanti. Sul tema si vedano: PARODI, SELLAROLI, *Sistema penale e intelligenza artificiale: molte esperienze e qualche equivoco*, in *Dir. pen. cont.* (Riv. Trim.), 6, 2019, pp. 55-59; SAPONARO, *Le nuove frontiere tecnologiche dell'individuazione personale*, in *Arch. Pen.*, 1, 2022.

³ È d'obbligo il richiamo a: QUATTROCOLO, *Artificial Intelligence, Computational Modelling and*

tematiche quali la prevedibilità dell'esito giudiziario, il giudizio di accertamento della pericolosità sociale, di commisurazione della pena e quello relativo alla concessione di benefici tra cui, ad esempio, la sospensione condizionale.

L'indagine che si intende condurre verterà specificamente su tale ultimo tema, allo scopo di verificare se (e, in caso di risposta affermativa, entro quali limiti) il ricorso a tale ultima frontiera tecnologica sia realizzabile e realmente auspicabile.

2. Tecnologia e giustizia penale: dalle banche dati giuridiche ai *software* intelligenti. Cosa si intende per giustizia predittiva?

L'utilizzo della tecnologia per coadiuvare gli operatori del diritto non rappresenta un fenomeno nuovo.

Ben prima dello sviluppo dell'intelligenza artificiale, si sono diffuse raccolte digitali di giurisprudenza in grado di reperire automaticamente precedenti giurisprudenziali sulla base delle parole chiave inserite, allo scopo di agevolare il lavoro tanto del giudice quanto delle altre parti del processo, le quali ricorrono alle banche dati allo scopo di conoscere gli orientamenti prevalenti e, dunque, elaborare un'adeguata strategia processuale⁴.

Attualmente, tuttavia, si fa strada la possibilità di una più incisiva interazione tra tecnologia e giustizia, resa possibile dalla diffusione dell'IA che ha permesso di ottenere un consistente miglioramento nella resa dei c.d. sistemi informatici esperti (SIE)⁵.

Criminal Proceedings. A Framework for A European Legal Discussion, Springer, Cham, 2020. Per un approfondimento sul punto si vedano anche: GIALUZ, *Intelligenza artificiale e diritti fondamentali in ambito probatorio*, in DI ARGENTINE (a cura di), *Giurisdizione penale, intelligenza artificiale ed etica del giudizio*, cit., pp. 51 e ss.; RICCIO, *Ragionando su intelligenza artificiale e processo penale*, in www.archiviopenale.it, pp. 7-10; PARODI, SELLAROLI, *Sistema penale e intelligenza artificiale*, cit., pp. 61-63; SAPONARO, *Le nuove frontiere tecnologiche*, cit.

⁴ Non è questa la sede per affrontare un problema tanto invisibile quanto serio, come quello della massimazione non soltanto spesso non fedele al contenuto delle sentenze ma soprattutto parziale, come tale potenzialmente incapace di rispecchiare la realtà e, dunque, forviante per le parti, per i giudici e per il sistema, atteso che in tal modo si spinge verso una stabilizzazione del precedente solo illusoria o, peggio ancora, errata nei contenuti. Invero, il carattere necessariamente parziale dei precedenti massimati – che, però, allo stato sono gli unici facilmente reperibili – può indurre a ritenere “prevalenti” orientamenti che in realtà sono molto meno radicati di quanto sembra.

⁵ I sistemi informatici esperti sono programmi che aspirano a consentire la risoluzione di problemi specifici attraverso il ricorso alla logica matematica. Per un inquadramento in chiave storica del fenomeno, si veda: VITERBO, CODIGNOLA, *I 70 anni del manifesto dell'intelligenza artificiale*, in *Dir. Inf.*, 4-5, 2007, pp. 725 e ss.; ID., *L'intelligenza artificiale e le sue origini culturali*, in *Giur. It.*, 7, 2014, pp. 1541 e ss.; SIMON, *Causalità, razionalità, organizzazione*, Bologna, 1985; KAPLAN, *Intelligenza artificiale. Guida al prossimo futuro*, Roma, 2017, pp. 37 e ss.

Poiché non vi è unanimità di vedute in ordine alla definizione di IA all'interno della comunità scientifica, si cercherà di spiegare sommariamente il modo attraverso cui tali tecnologie operano.

I “primi” tipi di SIE - che hanno rappresentato il punto di partenza per lo sviluppo dei moderni sistemi di IA - partendo da una “base di conoscenza” rappresentata da un insieme di fatti, regole e relazioni riguardanti uno specifico settore di interesse tradotti in linguaggio informatico, attraverso un insieme di operazioni matematiche (algoritmi), elaboravano le informazioni restituendo degli *output* che soddisfacevano un preassegnato insieme di requisiti.

Prima della diffusione dell'IA, l'efficacia e l'attendibilità dei risultati raggiunti dipendeva, per un verso, dalla qualità e dalla completezza della “base di conoscenza”, per altro, dalla competenza e dalla disponibilità degli ingegneri informatici ad immettere nel sistema i nuovi dati raccolti e a correggere eventuali errori di programmazione degli algoritmi⁶.

La moltiplicazione delle fonti in grado di rilevare i dati (che costituiscono la “base di conoscenza” dei SIE) e la diffusione di sistemi di *cloud* (che ne consentono l'elaborazione e l'archiviazione in rete) hanno gradualmente permesso di migliorare l'affidabilità delle correlazioni individuate dagli elaboratori, le quali, come noto, dipendono dalla quantità di dati a disposizione del sistema (oltre che dalla loro qualità). Parallelamente, sono stati sviluppati sistemi di *machine learning* che consentono alle macchine di accedere autonomamente ai dati conoscitivi e di processarli con la conseguenza che, oggi, queste sono in grado di individuare correlazioni e di dedurre regole senza che il programmatore sia di volta in volta chiamato a dover aggiornare i dati e le formule matematiche degli algoritmi⁷.

In sostanza, i moderni sistemi funzionano in virtù di uno schema analogico basato su associazioni deboli di dati, che in un certo senso riproduce il modo di apprendere dell'uomo e che spesso viene realizzato tramite l'impiego di reti neurali artificiali, cioè programmi ispirati da ipotetici principi organizzativi del cervello umano (c.d. *deep learning*)⁸.

⁶ Cfr. in tal senso: KAPLAN, *Intelligenza artificiale*, cit., pp. 50-51.

⁷ Sul funzionamento dei sistemi di *machine learning* si vedano: KAPLAN, op. cit., pp. 56 e ss. e MAGRO, *Decisione umana e decisione robotica. Un'ipotesi di responsabilità da decisione robotica*, in *La legislazione penale*, 10 maggio 2020, p. 2. Per un maggiore approfondimento: CARROZZA, ODDO, ORIVETO, DI MININ, MONTEMAGNI, *AI: profili tecnologici. Automazione e Autonomia: dalla definizione alle possibili applicazioni dell'Intelligenza Artificiale*, in *BioLaw Journal – Rivista di BioDiritto*, 3, 2019, p. 243 i quali precisano che gli algoritmi e i sistemi di dati alla base del *machine learning* si dividono in due categorie: gli algoritmi *supervised* nei quali il fattore umano entra in gioco fornendo una interpretazione nei dati o nelle regole dell'algoritmo di addestramento che influenza l'*output* finale e quelli *unsupervised* nei quali, invece, i dati e le correlazioni sono estratti in maniera automatica dal sistema che poi procede a processarli.

⁸ Come è efficacemente osservato da KAPLAN, *Intelligenza artificiale*, cit., p. 62 poiché il modo in cui effettivamente funziona il cervello umano non è del tutto pacifico, è bene definire i programmi informatici che funzionano sulla base di reti neurali artificiali come «imitatori incredibilmente

Lo sviluppo di questa nuova forma di tecnologia ha determinato un mutamento di paradigma. In altri termini, l'attenzione degli studiosi si è spostata dai tradizionali SIE verso i c.d. sistemi di autoapprendimento in quanto questi consentono (anche) «ai “non esperti” di utilizzare modelli e tecniche di apprendimento automatico senza essere competenti in quel campo di conoscenza, e senza neppure fornire alla macchina i dati necessari per elaborare il modello di automazione di un processo»⁹.

Con specifico riferimento alla decisione giudiziaria, l'utilizzo di *software* intelligenti fondati sui c.d. algoritmi di ragionamento capaci di trarre conclusioni coerenti con le premesse (cioè i dati inseriti nel sistema), consentirebbe di elaborare sia strumenti predittivi a fini decisori, sia strumenti di predizione decisoria.

Mentre i primi, in quanto dovrebbero fornire elementi su cui basare una decisione, vedono quali destinatari finali i giudici, i secondi, mirando a determinare la prevedibilità delle future decisioni giudiziarie su casi simili, si rivolgono in prima battuta alle parti del processo¹⁰.

Con l'espressione “giustizia predittiva” si intende «la capacità attribuita alle macchine di mettere in moto rapidamente in un linguaggio naturale il diritto pertinente per trattare un caso, di contestualizzarlo secondo specifiche caratteristiche e di anticipare la probabilità delle decisioni che potrebbero essere prese»¹¹. Pertanto, valorizzando l'ultimo inciso della definizione in parola, si dovrebbe concludere nel senso di riferire l'espressione ai soli strumenti di predizione decisoria (e non anche a quelli predittivi a fini decisori) atteso che solo i primi mirano a determinare la prevedibilità delle future decisioni su casi simili.

Tuttavia, è facile osservare che la distinzione di cui sopra ha una portata labile in quanto la diffusione di strumenti di predizione decisoria ben potrebbe finire per influenzare la decisione del giudice, sicché il termine giustizia predittiva viene spesso inteso in senso omnicomprensivo.

D'altra parte, lo stesso Autore da cui si è mutuata la definizione osserva che «presa alla lettera l'espressione potrebbe significare la sostituzione pura e semplice di un algoritmo alla giurisdizione composta da uomini e donne»¹².

talentuos» in quanto essi «“non imparano a fare qualcosa” (...), cioè non sviluppano una comprensione di principio delle relazioni e delle proprietà del loro mondo».

⁹ Così MAGRO, *Decisione umana e decisione robotica*, cit., p. 2.

¹⁰ La distinzione tra strumenti predittivi a fini decisori e strumenti di predizione decisoria è elaborata da KOSTORIS, *Predizione decisoria e diversione processuale*, in DI ARGENTINE (a cura di), *Giurisdizione penale, intelligenza artificiale ed etica del giudizio*, Milano, 2021, pp. 95.

¹¹ La definizione cui si è fatto riferimento è quella fornita da GARAPON, LASSÈGUE, *Justice digitale. Révolution graphique et rupture anthropologique*, Paris, 2018; trad. it., *La giustizia digitale. Determinismo tecnologico e libertà*, Bologna, 2021, p. 171.

¹² *Ibidem*.

3. I vantaggi del ricorso alla giustizia predittiva...

Il ricorso alla giustizia predittiva (*rectius*: l'utilizzo di strumenti di predizione decisoria sia per le parti del processo che per i giudici) si ritiene astrattamente in grado di apportare vantaggi al sistema penale; tuttavia, è innegabile che esso sia allo stesso tempo foriero di rischi.

Per ciò che concerne la posizione delle parti processuali, si osserva che i sistemi di predizione decisoria consentirebbero di addivenire ad una strategia processuale maggiormente efficace in quanto fondata non solo sulla bravura del Pubblico Ministero o del difensore, ma anche su un calcolo statistico legato alla possibilità di successo di quella strategia, il che, di riflesso, consentirebbe di ridurre il carico giudiziario, incentivando il ricorso a riti alternativi e, più in generale, a modelli di conclusione del procedimento diversi da quello ordinario¹³.

Parallelamente, dal punto di vista dell'organo giudicante, l'utilizzo di tali strumenti quale ausilio nell'emissione della decisione consentirebbe di accorciare i tempi del processo, promettendo di addivenire ad una realmente ragionevole durata dei giudizi.

Si tratta, in definitiva, di risultati pratici di notevolissima importanza, anche alla luce delle politiche europee che prevedono tra gli obiettivi da raggiungere – soprattutto per l'Italia – una maggiore efficienza del sistema giustizia da realizzare anche attraverso il ricorso alla digitalizzazione¹⁴.

Gli effetti positivi dell'uso dell'IA nell'ambito della decisione giudiziaria, però, non sarebbero di carattere esclusivamente pratico.

Si osserva, infatti, che la “*calcolabilità giuridico-penale*”, unitamente alla funzione degli strumenti di predizione decisoria, sarebbero progressivamente in

¹³ Al riguardo KAPLAN, *Intelligenza artificiale*, cit., p. 139 riporta un recente tentativo di applicazione della *machine learning* alla previsione dei verdetti della Corte Suprema statunitense che ha riportato un'affidabilità pari al 70%.

Parimenti degno di menzione è il progetto portato avanti dall'*University College of London* che nel 2016 ha sviluppato un algoritmo in grado di prevedere le decisioni della Corte europea dei diritti dell'uomo riportando un margine di successo che si aggira intorno al 79%. Il progetto può essere consultato al link <https://www.ucl.ac.uk/news/2016/oct/ai-predicts-outcomes-human-rights-trials>, mentre per la presentazione della ricerca si veda: ALETRAS, TSARAPATSANIS, PREOTJUC-PIETRO, LAMPOS, *Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective*, in *PeerJ Computer Science*, 2016 reperibile su <https://doi.org/10.7717/peerj-cs.93>. In argomento si veda: COSTANZI, *La matematica del processo: oltre le colonne d'Ercole della giustizia penale*, in *Questione Giustizia*, 4, 2018, p. 186.

¹⁴ Il crescente interesse per l'efficienza della giustizia è confermato anche dal Piano Nazionale di Ripresa e Resilienza - il programma di ripartizione delle risorse dei fondi stanziati dall'Unione europea allo scopo di risollevarne le sorti dell'economia interna degli stati membri a seguito della crisi sanitaria dovuta alla diffusione del coronavirus – il quale, pur non occupandosi specificamente del ruolo dell'IA in ambito penale, prevede tra gli obiettivi da raggiungere quello di eliminare i tempi morti, puntando (anche) sulla digitalizzazione. In argomento si veda: TECCE, *PNRR e riforma della giustizia penale: prospettive e limiti*, in *Processo penale e giustizia*, 4, 2021, pp. 981 e ss.

grado di «sostituire la legge»¹⁵, la quale ormai da tempo non appare più in grado di conformare l'esito giudiziario, consentendo almeno di raggiungere una maggiore prevedibilità dello stesso¹⁶. Ciò, a sua volta, implementerebbe la capacità di autodeterminazione dei cittadini restituendo agli stessi una maggiore fiducia nei confronti dello Stato.

In altri termini, data la crisi in cui attualmente versa il principio di legalità, causata, per un verso, dalla sempre maggiore importanza acquisita dalle Corti (costituzionali, interne ed europee) nel processo di formazione del diritto e, per altro, dalla tendenza della giurisprudenza a «forzare i cancelli delle parole»¹⁷ attraverso un uso distorto della tecnica di interpretazione estensiva¹⁸, si è fatta strada l'idea che il ricorso alla c.d. «giustizia predittiva» - essendo in grado di garantire una stabilizzazione «debole» del precedente giurisprudenziale - potrebbe rappresentare un rimedio in grado di assicurare il recupero almeno parziale delle garanzie di cui il principio di legalità si fa portatore in materia penale¹⁹.

Parimenti, gli strumenti di predizione decisoria sarebbero capaci di migliorare l'affidabilità dei giudizi prognostici che l'organo decidente è chiamato ad emettere tutte le volte in cui assume rilievo una valutazione sulla pericolosità di un soggetto e garantirebbero – attraverso il ricorso ad indici di gravità oggettivi - una maggiore uniformità nella determinazione del trattamento sanzionatorio, consentendo altresì al giudice di liberarsi, sia pur parzialmente, dal senso di responsabilità collegato alla necessità di dover esercitare «poteri, i quali suppongono una capacità che l'uomo non possiede»²⁰.

Il ricorso alla giustizia predittiva sembrerebbe, infine, incoraggiato dai

¹⁵ SGUBBI, *Il diritto penale totale. Punire senza legge, senza verità, senza colpa. Venti tesi*, Bologna, 2019, p. 41.

Nello stesso senso, si veda: GARAPON, LASSÈGUE, *Justice digitale*, cit., p. 175 che afferma «mentre le banche dati permettevano di conoscere meglio il diritto, di commentarlo e perfino di criticarlo, il trattamento tramite i *big data*, a partire da sentenze effettivamente emesse, genera una nuova normatività».

¹⁶ Muovono da queste premesse: DI GIOVINE, *Il Judge-Bot e le sequenze giuridiche in materia penale (intelligenza artificiale e stabilizzazione giurisprudenziale)*, in *Cass. Pen.*, 3, 2020, pp. 951 e ss.; CATERINI, *Il giudice penale robot*, in *La legislazione penale*, 19 dicembre 2020; MANES, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in *Discrimen*, 15 maggio 2020.

¹⁷ Cfr. IRTI, *I cancelli delle parole*, Napoli, 2015.

¹⁸ In argomento si veda MANNA, *Il lato oscuro del diritto penale*, Pisa, 2017, pp. 33 e ss. e 44 e ss.

¹⁹ Sull'importanza della prevedibilità del diritto e sulla necessità di evitare decisioni "a sorpresa", le quali imporrebbero quantomeno un obbligo di prendere in considerazione il precedente anche nei sistemi di *civil law* (nonostante l'assenza di un obbligo formale in tal senso) si veda: BIFULCO, *Il giudice è soggetto soltanto al «diritto»*. *Contributo allo studio dell'art. 101, comma 2 della Costituzione italiana*, Napoli, 2008, pp. 43 e ss.

²⁰ L'espressione è ripresa da CARNELUTTI, *La pena dell'ergastolo è costituzionale?*, in *Riv. dir. proc.*, 1, 1956, pp. 1-6 (spec. p. 4). Nello scritto l'Illustre giurista censura la funzione retributiva della pena affermando che «il dosaggio del *malum passionis* in anni, mesi e giorni secondo la gravità di un imponderabile *malum actionis* è quasi sempre meno serio del calcolo di Shylock quando pretendeva di valutare in una libbra di carne l'inadempimento del suo debitore».

numerosi studi di psicologia cognitiva che hanno dimostrato come le decisioni umane - anche quelle che riteniamo essere state prese nel modo più razionale possibile- sono per lo più il frutto di un uso inconsapevole (ma preponderante) di processi mentali automatici, intuitivi ed associativi spesso viziati da errori cognitivi (c.d. *biases*)²¹.

Se a tanto si aggiunge che l'attitudine ad incorrere nelle trappole cognitive è destinata ad aumentare allorché un'attività lavorativa di tipo intellettuale si protrae ininterrottamente per un certo lasso di tempo a causa dell'affaticamento delle aree cognitive coinvolte nel pensiero razionale²², si comprende facilmente l'interesse che l'utilizzo dell'IA nella fase decisoria suscita tra gli studiosi del diritto.

Tali strumenti, in definitiva, non opererebbero in modo così diverso dal cervello umano ma, a differenza di questo, avrebbero il pregio di non emettere decisioni condizionate da *biases* e non risentirebbero neppure della stanchezza mentale²³.

4. ... ed i rischi legati al suo utilizzo

Se il discorso in parola si arrestasse a quanto fin ora affermato, si dovrebbe concludere nel senso di incoraggiare l'utilizzo dei sistemi di IA in fase decisoria in quanto questi si presterebbero ad iniettare una certa dose di virtuosismo nel sistema penale.

Tuttavia, mentre è innegabile che il ricorso alla giustizia predittiva consentirebbe di ridurre notevolmente i tempi del processo, non è altrettanto vero che questa sarebbe sempre in grado di restituire decisioni migliori di quelle prese interamente dai giudici in carne ed ossa.

Il primo limite degli strumenti automatici di ausilio alla decisione giudiziaria è già emerso nell'esperienza giuridica d'oltreoceano, ove l'ormai notissimo caso

²¹ Per un approfondimento sul modo in cui le trappole cognitive sono destinate ad influenzare il ragionamento giuridico, si rinvia a: FORZA, MENEGON, RUMIATI, *Il giudice emotivo. La decisione tra ragione ed emozione*, Bologna, 2017.

²² *ibidem*, pp. 123 e ss. che riportano uno studio condotto da alcuni ricercatori della Columbia University in collaborazione con l'Università di Tel Aviv, all'esito del quale è emerso che la tendenza dei giudici ad emettere decisioni favorevoli è maggiore all'inizio della giornata lavorativa, mentre scema via via che il tempo trascorre, salvo poi ritornare all'apice a seguito di una pausa.

Si veda anche NIEVA-FENOLL, *Inteligencia artificial y proceso judicial*, trad. it., *Intelligenza artificiale e processo*, Torino, 2019, pp. 32 e ss.

²³ In argomento si veda DI GIOVINE, *Il Judge-Bot e le sequenze giuridiche in materia penale*, cit., p. 956 che, alludendo allo studio riportato nella nota 22), afferma che se davvero il giudice-Robot rappresentasse una garanzia rispetto ad errori e *biases* e se fosse vero che la giustizia dipende da ciò che i «giudici mangiano (o non mangiano) acquisterebbe senso ricorrere all'IA, perché diverrebbe ragionevole supporre che in tal modo si potrebbe assicurare una giustizia migliore».

*State v. Loomis*²⁴ ha dimostrato che non è corretto affermare che il ricorso ad una giustizia algoritmica elimina (o almeno riduce) il rischio di una decisione viziata da fattori psicologici, politici e culturali, in quanto l'utilizzo di tali sistemi si limita semplicemente a spostare il problema in una fase cronologicamente antecedente - quella della programmazione - e, parallelamente, solleva quesiti giuridici nuovi.

Nella vicenda menzionata - ove l'imputato, dopo essere stato condannato ad una pena che reputava sproporzionata per eccesso rispetto all'entità del reato commesso, aveva presentato (senza successo) ricorso alla Suprema Corte del Wisconsin - veniva contestato l'utilizzo del sistema COMPAS (un *software* che valuta il rischio "statico" di recidivismo) quale ausilio nell'attività di commisurazione della pena²⁵.

Nello specifico, la difesa lamentava l'attitudine del sistema ad emettere *output* viziati da pregiudizi basati sul genere e sulla razza²⁶, nonché l'opacità del suo funzionamento.

Senza scendere nel merito della decisione, in questa sede preme evidenziare due aspetti.

Il primo - già parzialmente anticipato nell'*incipit* del paragrafo - è che le decisioni algoritmiche possono soffrire degli stessi pregiudizi che talvolta inficiano le decisioni umane e che sono veicolati (spesso involontariamente) all'interno del ragionamento automatizzato dai programmatori, i quali sono chiamati a creare il contesto all'interno del quale il *deep learning* opera, consentendo alle macchine di auto-apprendere²⁷.

Il secondo, ben più complesso, attiene all'opacità del processo decisionale (*c.d. black box*) ed è destinato a ripercuotersi sulle garanzie del giusto processo.

²⁴ *State v. Loomis*, 881, NW 2d 749 (Wis 2016). Il testo della sentenza è reperibile in calce al contributo di CARRIER, *Se l'amicus curiae è un algoritmo: il chiacchierato caso Loomis alla Corte Suprema del Wisconsin*, in *Giurisprudenza Penale Web*, 2019, 4.

²⁵ Tra i numerosi contributi che fanno riferimento al caso *Loomis*, si veda: BASILE, *Intelligenza artificiale e diritto penale*, cit., pp. 21 e ss.; QUATTROCOLO, *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale "predittiva"*, in *Cass. pen.*, 4, 2019, pp. 1748 e ss.; ID., *Equo processo penale e sfide della società algoritmica*, in D'ALOIA (a cura di), *Intelligenza artificiale e diritto. Come regolare un mondo nuovo*, Milano, 2020, pp. 267 e ss., spec. pp. 277 e ss.

²⁶ Nello specifico, una ricerca condotta nel 2016 su incarico dell'ONG ProPublica avrebbe dimostrato che il sistema COMPAS sottostima il rischio di recidiva degli imputati bianchi e sovrastima quello degli imputati di colore (cfr. ANGWIN ET ALII, *Machine bias*, in <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>, 23 maggio 2016). A seguito della pubblicazione dei risultati della ricerca condotta dai ricercatori di ProPublica, un team di ricercatori di COMPAS ha effettuato un "controstudio" volto a criticare le conclusioni cui era giunto il precedente (FLORES, BECHTEL, LOWENKAMP, *False Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks"*, in *Federal Probation*, vol. 80, n. 2, 2016, pp. 38 e ss.

²⁷ Così CASONATO, *Per una intelligenza artificiale costituzionalmente orientata*, in D'ALOIA (a cura di), *Intelligenza artificiale e diritto*, cit., pp. 131 e ss., spec. p. 145

Da questo punto di vista, assume anzitutto rilievo la possibilità che i *software* utilizzati siano coperti da segreto commerciale e che, pertanto, sia impossibile accedere al codice sorgente che governa l'algoritmo, con la conseguenza di precludere *ab origine* la possibilità di conoscere il metodo di elaborazione dei dati effettuato dalla macchina e, dunque, di poterlo confutare²⁸.

È quanto accaduto nel caso *Loomis*, ove la difesa lamentava l'assenza di spiegazioni sui criteri con cui erano stati determinati i punteggi di rischio ed i fattori presi in considerazione, ponendo in tal modo a fondamento della decisione elementi sottratti alla *discovery*²⁹.

Come si vedrà nel paragrafo che segue, in Italia è possibile fin da ora escludere l'eventualità che si verifichi un caso analogo a quello insorto nel caso "guida" oltreoceano (cioè che si ricorra a *software* predittivi coperti da segreto commerciale)³⁰.

Tuttavia, il problema dell'opacità del processo decisionale effettuato dagli algoritmi è destinato comunque ad assumere rilevanza in quanto l'accessibilità al codice sorgente, agli *inputs* e agli *outputs* del *software* non garantisce né al giudice che si fa coadiuvare da tali sistemi, né alla parte che è destinata a subirne gli effetti la possibilità di trarre autonomamente elementi significativi e comprensibili dal sistema.

Tale aspetto, peraltro, se fronteggiato mediante il ricorso a strumenti processuali, quali la consulenza di parte, finirebbe semplicemente per spostare i termini della questione dalla necessità di conoscere i meccanismi di funzionamento del *software* per poterli confutare alla materiale e concreta possibilità di farlo.

Allo stato, infatti, i sistemi intelligenti sono comprensibili solo da chi li ha creati e li gestisce³¹ ed è dunque ragionevole presumere che, in caso di introduzione degli stessi, sarà la parte pubblica (cioè l'accusa) ad avere accesso alla scienza e alle tecnologie migliori, salvo che il giudice non ritenga opportuno disporre una perizia.

L'introduzione di sistemi di giustizia predittiva, pertanto, potrebbe acuire lo squilibrio conoscitivo tra le parti del processo, negando di fatto all'imputato la possibilità di presentare i propri argomenti in condizioni di parità rispetto all'altra parte processuale³².

²⁸ Cfr. QUATTROCOLO, *Equo processo penale*, cit., pp. 278 e ss. La necessità dell'impiego da parte dell'amministrazione di utilizzare *software* c.d. *open data* in luogo di quelli "proprietary" allo scopo di garantire la "conoscibilità" dell'algoritmo è segnalata anche da DONATI, *Intelligenza artificiale e giustizia*, in D'ALLOIA (a cura di), *Intelligenza artificiale e diritto*, pp. 237 e ss. (spec. pp. 249-250).

²⁹ ¶6 *State v. Loomis*, 881, NW 2d 749 (Wis 2016), cit.

³⁰ *Infra* par. 5.

³¹ Per un approfondimento sul problema dell'opacità dei sistemi di Intelligenza artificiale e sui pericoli che tale aspetto comporta, si rinvia a: FIORIGLIO, *La società algoritmica fra opacità e spiegabilità: profili informatico-giuridici*, in *Ars interpretandi*, 1, 2021, pp. 53 e ss. (spec. p. 58).

³² Si sofferma sulle conseguenze che il problema della inconfutabilità del dato elaborato dall'IA può avere sul principio di parità delle armi e, più in generale, del giusto processo:

Ma vi è di più. La complessità dei moderni sistemi di IA, unitamente alla capacità di cui questi sono dotati di accedere autonomamente ai dati conoscitivi e di processarli, rende spesso difficile anche a chi li detiene la comprensione delle ragioni che hanno indotto la macchina ad emettere un determinato *output*, rendendo in concreto imprevedibili e difficilmente spiegabili le decisioni da questa fornite³³.

In altri termini, la complessità del linguaggio informatico proprio degli algoritmi genera quello che è stato definito «un analfabetismo di ritorno», atteso che relega il giurista a un mero fruitore ignorante di un linguaggio di cui gli è preclusa la possibilità di apprendere la grammatica e la sintassi³⁴.

La difficoltà di confutare i risultati cui perviene l'algoritmo porta con sé anche il rischio di aggravare il fenomeno del c.d. "effetto gregge"³⁵.

Già attualmente, infatti, soprattutto tra i giudici di merito, vi è la tendenza a replicare le decisioni prese in casi analoghi sulla scorta di una valutazione che tiene conto, non solo dell'autorevolezza e della correttezza logico argomentativa del precedente reperito, ma anche del numero di sentenze che recepiscono – in modo più o meno pertinente rispetto al caso concreto – il medesimo principio di diritto. Ciò è dovuto ad una serie di concause: *in primis*, l'elevato carico di lavoro presente in primo grado ed in appello rende impossibile uno studio approfondito di tutti i casi da decidere, con la conseguenza che l'attenzione è destinata a concentrarsi sui processi ritenuti più "difficili"³⁶. In secondo luogo, la tendenza a seguire le decisioni già prese dalla maggioranza è incentivata dalla parziale "deresponsabilizzazione" che questa comporta, atteso che il magistrato il quale aderisce (a volte acriticamente) agli orientamenti maggioritari non sarà etichettato come la "pecora nera"³⁷ dai suoi colleghi (i quali, peraltro, potrebbero essere chiamati a valutarne l'operato) e, allo stesso tempo, potrà condividere con l'intera categoria il peso delle (sempre più frequenti) critiche provenienti dall'opinione pubblica.

Nel contesto appena delineato, è ragionevole desumere che l'introduzione di strumenti di predizione decisoria quale ausilio sia per i giudici che per le parti del processo spinga inconsapevolmente verso un conformismo generalizzato³⁸, oltre

QUATTROCOLO, *Equo processo penale*, cit., pp. 273 e ss.

³³ Si occupa del problema della *black box* anche: CASONATO, *Intelligenza artificiale e giustizia: potenzialità e rischi*, in *DPCE Online*, 3, 2020, pp. 3369 e ss. (spec. pp. 3376 e ss.).

³⁴ GARAPON, LASSEGUE, *Justice digitale*, cit., p. 202.

³⁵ *Ibidem*, pp. 184 e ss. Si occupa del fenomeno dell'effetto gregge e dell'amplificazione che lo stesso avrebbe qualora venissero introdotti strumenti di IA quale ausilio all'attività delle parti del processo penale anche: DI GIOVINE, *Il Judge-Bot*, cit., pp. 959 e ss.

³⁶ Prendendo le mosse (anche) da questa premessa DI GIOVINE, *Il Judge-Bot*, cit. introduce la distinzione tra casi "facili" e "difficili" e si interroga in ordine alla possibilità di utilizzare l'IA in fase decisoria come ausilio per la decisione di casi "facili".

³⁷ GARAPON, LASSEGUE, *Justice digitale*, cit., p. 214.

³⁸ *Ibidem* p. 176, ove l'Autore afferma che «la logica predittiva genera un sapere sempre più

che verso una acritica uniformità decisionale, a fronte di mutamenti sociali che la giurisprudenza non si mostrerebbe più in grado di intercettare e a cui, quindi, cesserebbe di dare risposta³⁹.

Non si può non tener conto, d'altronde, del concreto pericolo che si diffonda l'idea che la macchina sia infallibile per il solo fatto che è capace di processare un numero infinitamente elevato di dati e trarre dagli stessi correlazioni statistiche attendibili (anche se spesso difficilmente spiegabili) dimenticandosi, al contempo, di avere a che fare con strumenti che non forniscono un risultato in grado di tener conto delle specificità del caso concreto⁴⁰.

Peraltro, se tali timori dovessero effettivamente inverarsi e se si ritenesse davvero opportuno sacrificare l'argomentazione logico-giuridica in favore della prevedibilità del diritto e (soprattutto) della ragionevole durata del processo, non vi sarebbero ostacoli a passare dagli strumenti di predizione decisoria ai sistemi predittivi a fini decisori in senso stretto.

In altri termini, se si accettasse l'idea di sostituire il libero convincimento del giudice - che attualmente rappresenta il fulcro dell'accertamento della verità processuale - con il risultato delle correlazioni statistiche di impronte e di segni che stanno alla base del ragionamento informatico, non vi sarebbero motivi per non delegare la giustizia alle macchine. La scrittura digitale della giustizia, inoltre, comporterebbe - quale logico corollario - una sempre maggiore "disintermediazione del processo", atteso che "il rituale giudiziario" che caratterizza l'attuale processo penale - oltre ad allungare i tempi e ad aumentare i costi - sarebbe del tutto inutile per una giustizia informatizzata⁴¹.

5. Le politiche adottate in ambito comunitario: il ridimensionamento dei rischi legati all'IA in fase decisoria

Le perplessità evidenziate nel paragrafo precedente possono essere parzialmente ridimensionate alla luce della politica adottata in materia a livello comunitario.

L'Unione europea, ormai da tempo, si sta attivando affinché sia tracciata una "via etica" per lo sviluppo e l'utilizzo dei sistemi di IA volta a sfruttare al meglio

orizzontale. (...). È il volume delle correlazioni - e non più il ragionamento o l'autorità dei principi che fa il diritto».

³⁹ In tal senso si veda LUCIANI, *La decisione giudiziaria robotica*, in rivista AIC, 3, 2018, pp. 872 e ss. (spec. p. 887) secondo cui il vincolo «alla giurisprudenza pregressa impedisce l'evoluzione degli indirizzi giurisprudenziali e preclude al diritto di esercitare la sua funzione primaria (rispondere a bisogni umani regolando umani rapporti corrispondentemente alle esigenze sociali del momento storico)».

⁴⁰ Non a caso vi è chi in dottrina afferma che «l'applicazione dell'AI all'interpretazione giuridica» sia «semplicemente inimmaginabile». Così si esprime ZACCARIA, *Mutazioni del diritto: innovazione tecnologica e applicazioni predittive*, in *Ars Interpretandi*, 1, 2021, pp. 29 e ss. (spec. pag. 42).

⁴¹ Sulla funzione del ritualismo giudiziario si veda: GARAPON, LASSÈGUE, *Justice digitale*, cit., pp. 137 e ss.

tali tecnologie scongiurando il rischio che ciò vada a detrimento dei diritti fondamentali⁴².

Per quanto attiene all'apporto che tali strumenti potrebbero fornire in fase decisoria, occorre premettere che, allo stato, non vi è ancora una disciplina compiuta. Tuttavia, taluni principi utili possono essere ricavati dal Regolamento sulla protezione dei dati personali (GDPR)⁴³ e dalla Carta etica Europea sull'utilizzo dell'Intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi⁴⁴. Quest'ultimo è uno strumento di *soft law* sviluppato dalla Commissione Europea per l'efficienza della giustizia (CEPEJ) che ha visto la luce nel 2018 e che continua a rappresentare a livello europeo e mondiale il più completo testo specificatamente dedicato al tema.

Oltre a questi strumenti, merita di essere menzionata la direttiva sulla protezione dei dati nelle attività di polizia e giudiziaria⁴⁵ nonché lo studio su algoritmi e i diritti umani realizzato dal Consiglio d'Europa nel marzo del 2018⁴⁶.

Dalla lettura congiunta del GDPR e della Carta etica può anzitutto escludersi il rischio più grave tra quelli paventati nel paragrafo precedente, cioè che si arrivi ad una giustizia interamente automatizzata.

Il primo ed il quinto principio enunciato dalla Carta etica, infatti, sottolineano l'importanza che l'utilizzo di strumenti di IA non escluda il decisore umano e non inquina l'*iter* logico-argomentativo del giudice, rispetto al quale le nuove tecnologie devono rappresentare esclusivamente un supporto.

Sulla stessa lunghezza d'onda si colloca la normativa contenuta nel GDPR (le cui previsioni sono direttamente applicabili negli Stati membri) che, pur non essendo un testo normativo specificatamente dedicato alla materia oggetto di trattazione, cristallizza alcune regole fondamentali⁴⁷, tra le quali il divieto, stabilito all'art. 22 GDPR, alla conduzione di un processo decisionale completamente automatizzato – compresa la profilazione – che produca effetti sulla sfera giuridica dell'interessato o incida significativamente in modo analogo sulla sua persona.

⁴² Sulle politiche assunte in ambito comunitario per lo sviluppo dell'IA, tra tutti, si vedano: ZANICHELLI, *Ecosistemi, opacità, autonomia: le sfide dell'intelligenza artificiale in alcune proposte recenti della Commissione europea*, in D'ALOIA (a cura di), *Intelligenza artificiale e diritto*, cit., pp. 67 e ss.; DE FELICE, *Intelligenza artificiale e processi decisionali automatizzati: GDPR ed ethics by design come avamposto per la tutela dei diritti umani*, *ibidem*, pp. 415 e ss.

⁴³ Reg. 679/2016.

⁴⁴ Il documento è disponibile al seguente indirizzo: <https://rm.coe.int/carta-etica-europea-sull-utilizzo-dell-intelligenza-artificiale-nei-si/1680993348>. Per un commento, si veda: 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*, 18 dicembre 2018.

⁴⁵ Direttiva (UE) 2016/680 del 27 aprile 2016.

⁴⁶ Cfr. Studio del Consiglio d'Europa, *Algorithms and Human Rights*, DGI(2017), 12 marzo 2018 il cui testo è disponibile al sito <https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5>.

⁴⁷ In argomento si veda: DE FELICE, *Intelligenza artificiale e processi decisionali automatizzati*, cit., pp. 416-419.

Come è stato osservato, affinché tale divieto trovi applicazione, non è necessario che anche la decisione sia automatizzata, essendo, per contro, sufficiente che ad essere automatizzati siano le prove preparatorie e i giudizi discrezionali⁴⁸.

Da tanto deriva il ripudio in materia penale non solo di decisioni interamente meccaniche, ma anche di una totale automazione dell'istruttoria dibattimentale, con la conseguenza di escludere la possibilità che si realizzi un allontanamento dal ritualismo giudiziario, il quale, lungi dall'essere un inutile spreco di tempo e di risorse, ha invece il compito di "avvicinare" il giudice alle parti (e le parti al giudice) consentendo di conservare la funzione (anche) sociale che tradizionalmente la giustizia è chiamata a svolgere⁴⁹.

Quanto appena affermato, inoltre, non risulta smentito neppure dalle deroghe al divieto di un processo decisionale completamente automatizzato contenute all'interno dell'art. 22 par. 2 GDPR, atteso che lo stesso regolamento al paragrafo successivo sancisce il c.d. principio dell'*under user control*, cioè prescrive che sia assicurata in ogni caso la contestabilità della decisione nonché che sia assicurato all'interessato, almeno in *extrema ratio*, il filtro dell'uomo⁵⁰.

La prospettiva assunta a livello comunitario, inoltre, pone al centro il principio di trasparenza nella duplice accezione che coinvolge sia l'interesse a che il metodo di funzionamento dei sistemi intelligenti impiegati in ambito giudiziario non sia coperto da segreto commerciale, sia il versante relativo alla necessità di superare l'opacità intrinseca dei sistemi di IA.

Sotto il primo profilo, la Carta etica sottolinea la necessità di trovare un equilibrio tra proprietà intellettuale e necessità di conoscere il metodo di funzionamento dei sistemi intelligenti impiegati in ambito giudiziario proponendo a tal fine l'introduzione di un sistema di certificazione⁵¹. Inoltre, l'art. 15 lett. h) GDPR stabilisce il diritto dell'interessato a conoscere l'esistenza di un processo decisionale automatizzato nonché di ottenere informazioni significative sulla logica utilizzata⁵².

Per ciò che attiene all'altro versante della trasparenza, invece, la prospettiva assunta in ambito comunitario è volta a favorire il raggiungimento di una *explainable* IA. Più nello specifico, il terzo principio enunciato dalla Carta etica pone l'accento sulla necessità che sia garantito un sistema di controllabilità *ex post* del processo decisionale utilizzato dalla macchina nonché di tracciamento dei

⁴⁸ *Ibidem*, p. 417.

⁴⁹ In questo senso: GARAPON, LASSÈGUE, *Justice digitale*, cit., pp. 152-153.

⁵⁰ Cfr. DE FELICE, *Intelligenza artificiale e processi decisionali automatizzati*, cit., p. 418 per il quale le ulteriori garanzie sancite dall'art. 22 par. 3 GDPR risultano fondamentali per contrastare la delega incondizionata al determinismo della tecnologia. In senso contrario, tuttavia, si veda ZANICHELLI, *Ecosistemi, opacità, autonomia*, cit., p. 87 la quale invece afferma che «sembra mancare ad oggi una chiara opzione di fondo favorevole ad un'IA che dia supporto e assistenza all'essere umano e contraria a un'IA che tenda a sostituirlo».

⁵¹ Si veda il 4° principio della Carta etica, p. 11.

⁵² Cfr. anche appendice I alla Carta etica, p. 38.

dati processati, i quali sono alla base dell'*output* restituito.

Infine, la Carta etica sottolinea la necessità di creare le condizioni per armonizzare i dati allo scopo di gettare le basi affinché siano sviluppati sistemi di IA da impiegare in fase decisoria che non risentano di pregiudizi propri di chi li programma.

In definitiva, l'approccio "etico" che l'Unione europea ha assunto rispetto allo sviluppo e alla diffusione dei sistemi di IA consente di scongiurare gran parte dei rischi esposti nel paragrafo precedente e ciò anche perché è accompagnato da un atteggiamento estremamente cauto rispetto all'impiego di tali strumenti in ambito giudiziario.

Invero, dalla lettura dell'Introduzione e dell'appendice I alla Carta etica emerge il ripudio di un approccio basato sull'utilizzo dell'IA a tutti i costi, in special modo allorché entra in gioco la sanzione penale. Ciò, pertanto, consente di affermare che le regole appena enunciate non saranno relegate a mera dichiarazione di intenti, ma orienteranno in maniera effettiva lo sviluppo di sistemi di giustizia predittiva all'interno del processo (anche penale), eventualmente impedendone l'utilizzo laddove ciò dovesse comportare il rischio di discriminazioni basate su dati sensibili, dovesse mettere in discussione le garanzie previste in materia di equo processo ovvero si dovesse porre in contrasto con le funzioni della pena⁵³.

Quanto appena affermato sull'atteggiamento assunto dall'Unione europea rispetto all'impiego dell'IA trova conferma nella proposta di regolamento volta ad individuare delle regole armonizzate sull'Intelligenza Artificiale ed il cui testo sarà destinato ad integrare la normativa già esistente⁵⁴. Non si tratta di una proposta specificamente dedicata al tema dell'impiego dei sistemi di IA in ambito giudiziario; tuttavia essa permette di ricavare talune informazioni utili anche in tale settore.

In particolare, la proposta in esame adotta un approccio basato sul rischio, differenziando tra gli utilizzi dell'IA che creano un rischio inaccettabile, quelli che creano un rischio alto e quelli che creano un rischio basso o minimo.

All'interno di questa classificazione i sistemi di giustizia predittiva dovrebbero essere collocati nell'ambito dei sistemi ad alto rischio⁵⁵, sicché il loro utilizzo – pur non essendo astrattamente illecito - dovrà essere subordinato al rispetto di requisiti obbligatori nonché ad una valutazione di conformità.

È interessante notare che rispetto all'impiego dell'IA in fase decisoria, la proposta di regolamento – pur collocandosi nel solco del sentiero già tracciato dalla Carta etica del 2018 - sembra fare alcuni passi avanti in quanto, oltre a

⁵³ In tal senso si veda l'appendice I, pp. 34 e ss.

⁵⁴ Cfr. Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica di alcuni atti legislativi dell'Unione, 21 aprile 2021, COM(2021) 206 final reperibile al seguente *link*: <https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX%3A52021PC0206>.

⁵⁵ Cfr. Considerando 40) della proposta di Regolamento.

ribadire la necessità di un controllo umano significativo dei sistemi ad alto rischio, prescrive esplicitamente la necessità che si assicuri non solo la trasparenza sull'operato di sistemi di IA ma anche la possibilità per gli utenti⁵⁶ di poter interpretare gli *output* del sistema in modo adeguato⁵⁷.

Si tratta di una previsione importante, atteso che la sua declinazione in chiave processuale dovrebbe limitare il rischio che si realizzi quello squilibrio conoscitivo tra le parti che – come è stato evidenziato nel paragrafo precedente – andrebbe a violare il principio di parità delle armi e dell'effettività del diritto di difesa.

È bene aggiungere, per completezza, che un parziale ridimensionamento dei rischi connessi all'utilizzo dei sistemi di IA in ambito giudiziario non deriva esclusivamente dall'analisi della normativa e delle politiche adottate in ambito comunitario. Invero, alcuni spunti di riflessione potrebbero essere tratti anche dall'analisi del diritto interno – sia vigente che vivente – dal quale si desumono principi e garanzie che portano ad escludere o comunque a limitare l'utilizzo di tali strumenti nel caso in cui questi non dovessero essere in grado di conformarsi.

A tal proposito, una parte della dottrina ha osservato che «in un sistema come il nostro, dominato dal principio di sottoposizione del giudice solo alla legge (art. 101 comma 2 Cost.), né gli strumenti predittivi a fini decisori, né gli strumenti di predizione decisoria sembrano poter determinare effetti vincolanti»⁵⁸. Sulla stessa scia, i giudici del Consiglio di Stato – che per primi sono stati chiamati a confrontarsi con una decisione emessa sulla base delle risultanze fornite da un algoritmo (nella specie un provvedimento amministrativo) – hanno affermato che «l'utilizzo di procedure robotizzate non può essere motivo di elusione dei principi che conformano il nostro ordinamento»⁵⁹, sicché deve essere contemplata la possibilità che il giudice possa svolgere sul piano umano valutazioni e accertamenti fatti direttamente in via automatica.

Pertanto, affinché tali sistemi possano trovare ingresso nel nostro ordinamento, non basta che sia garantita la conoscibilità del codice sorgente, essendo, per contro, necessario che «la “formula tecnica” che di fatto rappresenta l'algoritmo sia corredata da spiegazioni che la traducono nella regola giuridica ad essa sottesa e che la rendono leggibile e comprensibile sia per i cittadini che per il giudice»⁶⁰.

D'altro canto, il mancato rispetto dei suddetti principi, specialmente con riguardo all'applicazione del diritto penale, che è per natura destinato a limitare

⁵⁶ Ai sensi dell'art. 3 par. 1 lett. 4) della proposta di Regolamento si considera “utente” «qualsiasi persona fisica o giuridica, autorità pubblica, agenzia o altro organismo che utilizza un sistema di IA sotto la sua autorità, tranne nel caso in cui il sistema di IA sia utilizzato nel corso di un'attività personale non professionale».

⁵⁷ Cfr. Considerando 47) e 48) della proposta.

⁵⁸ Così KOSTORIS, *Predizione decisoria e diversion processuale*, cit., p. 96.

⁵⁹ Consiglio di Stato, Sez. VI, 8 aprile 2019, n. 2270 par. 8.2).

⁶⁰ *Ibidem*, par. 8.3).

diritti fondamentali della persona, potrebbe porsi in contrasto con la giurisprudenza costituzionale nella parte in cui sarebbe idoneo ad escludere ogni apprezzamento della situazione da parte del giudice, in tal modo introducendo surrettizi automatismi sanzionatori svincolati da riscontri pratici attendibili e verificabili⁶¹.

6. Osservazioni conclusive

Dall'analisi condotta emerge la chiara intenzione sia a livello nazionale, sia a livello comunitario di preservare una "giustizia umana" anche a costo di limitare l'impiego della nuova tecnologia.

È altrettanto evidente, tuttavia, che qualora le barriere tecnologiche attualmente insite negli strumenti di *machine learning* dovessero essere abbattute in modo da garantire un'effettiva trasparenza e tracciabilità dei dati immessi nel sistema nonché una reale conoscibilità e controllabilità *ex post* del processo decisionale utilizzato dalla macchina non vi sarebbero ostacoli a livello normativo per escludere l'utilizzo degli strumenti di predizione decisoria quale ausilio per l'attività delle parti e dell'organo decidente, a condizione, però, che questo conservi la possibilità di discostarsi dall'*output* emesso dal sistema quando questo non appaia conforme rispetto alle specificità del caso concreto.

Intesa in questi termini, la "digitalizzazione" consentirebbe «ai giudici e a tutti gli attori della giustizia di prendere coscienza delle proprie "storture cognitive", senza cadere nell'illusione di una loro totale eliminazione»⁶².

La rapidità con cui le nuove tecnologie evolvono induce a ritenere che lo scenario appena descritto non sia così lontano dal divenire realtà. Pertanto, è opportuno iniziare a chiedersi se, quando ciò accadrà, l'impiego dell'IA in fase decisoria dovrà effettivamente essere incoraggiato o se, per contro, sussistano criticità più profonde rispetto ad una parziale digitalizzazione della giustizia penale.

A ben vedere, nonostante gli sforzi messi in atto per garantire un "controllo umano significativo" il vero ostacolo che, allo stato, appare insormontabile risiede nel fascino insito nell'algoritmo ed ha natura eminentemente pratica. Infatti, «una volta introdotto un sistema automatico di decisione all'interno di un

⁶¹ Il rilievo critico per il quale l'ineffettività del principio di non esclusività della decisione algoritmica si porrebbe in contrasto con la giurisprudenza costituzionale in materia di automatismi sanzionatori è avanzato da SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *BioLaw Journal - Rivista di biodiritto*, 1, 2019, pp. 63 e ss. (spec. p. 81). La Consulta, infatti, ha affermato che «l'automatismo che esclude ogni apprezzamento della situazione del giudice (...), imposta pur quando essa appaia inadatta in concreto, infrange l'equilibrio costituzionalmente necessario e viola esigenze essenziali di protezione dei diritti della persona». Cfr. *ex multis* C. Cost. 29 novembre 2004, n. 367.

⁶² GARAPON, LASSEGUE, *Justice digitale*, cit., p. 274.

processo decisionale umano, il sistema automatico tende, nel tempo, a catturare la decisione stessa» sia perché «solleva il decisore dal *burden of motivation*, dal peso dell'esame e della motivazione» sia perché «gli consente di “qualificare” la propria decisione con il crisma di scientificità (...) che oggi circonda la valutazione algoritmica»⁶³.

Non può cioè escludersi che sia proprio la responsabilità del decidere a garantire la bontà del sistema e che la tenuta della giustizia penale sia da attribuire – almeno parzialmente – alla consapevolezza propria di ciascun “giudice uomo” di essere chiamato ad esercitare un compito idoneo a produrre conseguenze spesso irreversibili per la vita dei suoi simili. Da questo punto di vista, l'introduzione delle macchine, essendo idonea a causare una parziale deresponsabilizzazione del decidente, potrebbe inconsapevolmente rendere il giudice meno sensibile al peso delle conseguenze che la sua attività comporta sulle vite altrui e, quindi, più facilmente incline ad adottare soluzioni rigorose anche in presenza di casi “dubbi”.

In altri termini, quantomeno con riferimento al momento decisionario in senso stretto (*i.e.* quello che attiene alla fase di accertamento della responsabilità penale, della sussistenza di eventuali circostanze, ecc.), è proprio l'accentuato rischio di un “effetto gregge”⁶⁴ a rendere tendenzialmente inaccettabile l'integrazione tra decisione umana e digitale.

Per non parlare del rischio di privare – attraverso il ricorso all'IA - il sistema giustizia di quel «senso del giusto» che «affonda nel *moral sense*» che è «l'esatto contrario della computazione per regole»⁶⁵ e che, forse nemmeno troppo alla lunga, finirebbe con il negargli anche quel residuo consenso sociale di cui ad oggi gode, producendo un effetto non giustificabile in nome di una (pur importante ma forse irrealizzabile) prevedibilità ed omogeneità dell'esito giudiziario né tantomeno per il perseguimento di esigenze deflative.

Un discorso parzialmente diverso può essere condotto con riferimento agli accertamenti che implicano una valutazione prognostica di carattere probabilistico avente ad oggetto la pericolosità sociale di un soggetto. È quanto normalmente accade in sede di applicazione di misure di sicurezza, di prevenzione, di applicazione delle misure cautelari (con riferimento alla valutazione relativa all'esigenza di evitare una recidiva *ex art. 274 lett. c) c.p.p.*), di concessione della sospensione condizionale della pena ovvero di misure alternative alla detenzione.

In tutti questi casi assume rilievo un giudizio caratterizzato da un elevato grado di predittività rispetto al quale l'assenza di una metodologia univoca e di criteri sufficientemente definiti a livello legislativo per condurre l'accertamento

⁶³Si veda: SIMONCINI, *L'algoritmo incostituzionale*, cit., p. 81.

⁶⁴V. *Infra* par. 4.

⁶⁵Così DI GIOVINE, *Dilemmi morali e diritto penale*, Bologna, 2022, p. 187.

È sostanzialmente dello stesso avviso anche GARAPON, LASSÈGUE, *Justice digitale*, cit., p. 278.

rischia di far prevalere il metodo intuitivo⁶⁶.

Peraltro, questa incertezza metodologica che induce all'intuizionismo fa sì che la giurisprudenza tenda a prediligere la misura – tra quelle applicabili – maggiormente afflittiva: in caso di errore, infatti, si tratterà di un falso positivo difficilmente verificabile poiché non sarà mai dato sapere se il soggetto, lasciato in libertà, si sarebbe o meno astenuto dal commettere altri reati.

In questo contesto viene quindi spontaneo chiedersi se il ricorso alle tecniche di giustizia predittiva – una volta superati i problemi tecnici attualmente esistenti e di cui si è ampiamente trattato nei paragrafi precedenti – possa giocare un ruolo in questo ambito.

Il problema più importante, ancora una volta, attiene alla necessità di scongiurare i rischi connessi all'utilizzo di logiche ciecamente "statistiche" che non tengono conto delle peculiarità della singola persona.

A tal fine vi è chi in dottrina ha prospettato la possibilità di un utilizzo degli algoritmi esclusivamente *in bonam partem* e la cui operatività andrebbe calata all'interno di un giudizio di tipo bifasico. In tal modo il giudizio algoritmico consentirebbe, infatti, di mitigare gli effetti della decisione in presenza di una limitata capacità a delinquere, favorendo altresì l'identificazione del trattamento più adatto (tra quelli meno afflittivi) alle caratteristiche del soggetto. Al contempo, però, la previsione di un giudizio bifasico consentirebbe in ogni caso di rispettare il diritto ad una decisione individualizzata⁶⁷.

Inteso in questi termini, l'apporto dell'IA in fase decisoria potrebbe prestarsi – sia pur con estrema cautela – a produrre effetti benefici in quanto idoneo a rendere maggiormente oggettivo un giudizio che, allo stato, è privo parametri certi e, parallelamente, ad incentivare la giurisprudenza a superare i *biases* di cui spesso cade vittima in questi settori e che la inducono ad assumere un atteggiamento particolarmente rigoroso anche a causa del peso collegato alla responsabilità di dover emettere delle decisioni che (a differenza di quanto accade con riferimento all'accertamento della responsabilità) sono per natura destinate a guardare al futuro.

Peraltro, se è vero che la decisione algoritmica è dotata di una forza persuasiva tale da catturare il decisore umano, l'utilizzo di un sistema in grado di restituire un esito consistente nella misura meno afflittiva tra quelle applicabili, unitamente alla previsione che il decisore umano conservi la possibilità di discostarsi dall'*output* emesso sulla scorta di una valutazione individualizzante e previa motivazione, si presterebbe a rendere il trattamento sanzionatorio *lato sensu* inteso maggiormente aderente alle finalità rieducative

⁶⁶ In questo senso: MAUGERI, *L'uso di algoritmi predittivi per accertare la pericolosità sociale: una sfida tra evidence based practices e tutela dei diritti fondamentali*, in *Arch. Pen.*, 1, 2021, p. 7. Si veda anche p. 9 ove l'Autrice sottolinea che gli algoritmi «progettano una procedura normativa che si sposta da un insieme di dati verso un *output* desiderato, escludendo le intuizioni soggettive e l'arbitrarietà del processo».

⁶⁷ L'ipotesi di studio è formulata da MAUGERI, op. cit., pp. 29 e ss.

fatte proprie dalla Costituzione.

Per quanto suggestivo, il discorso in parola è per il momento destinato a rimanere sul piano meramente teorico. Allo stato, infatti, permane quella opacità nei processi decisionali automatizzati che esclude la possibilità di un loro utilizzo in ambito giudiziario, pena il rischio di sovvertire buona parte delle garanzie processuali esistenti.

