Intelligenza Artificiale e decisione penale[*]

di Fabio Pinelli[**]

1. *I sistemi d'intelligenza artificiale: il profilo definitorio e la sua declinazione in ambito giuridico e penale*

La nozione d'intelligenza artificiale e la sua possibile declinazione in ambito giuridico scontano inevitabilmente un profilo d'indeterminatezza[1], coinvolgendo una miriade di strumenti e applicazioni di carattere tecnico-informatico, difficilmente accomunabili in un'unica definizione di sintesi.
Nella sua comunicazione dal titolo *"L'intelligenza artificiale per l'Europa"*, del 2018, la Commissione Europea ha inteso definirla nei termini di quei *"sistemi che mostrano un comportamento intelligente analizzando il proprio ambiente e compiendo azioni, con un certo grado di autonomia, per raggiungere specifici obiettivi. I sistemi basati sull'IA possono consistere solo in software che agiscono nel mondo virtuale (ad esempio assistenti vocali, software per l'analisi delle immagini, motori di ricerca, sistemi di riconoscimento vocale e facciale) oppure incorporare l'IA in sistemi hardware (per esempio in robot avanzati, auto a guida autonoma, droni o applicazioni dell'Internet delle cose)"*[2].
Come ricorda il recente *"Libro bianco sull'intelligenza artificiale – Un approccio europeo all'eccellenza e alla fiducia"*, adottato dalla Commissione Europea il 19 febbraio 2020, la definizione appena citata è stata poi ulteriormente perfezionata dal Gruppo di esperti ad alto livello nominato dalla Commissione stessa, nel giugno del 2018, *"per fornire consigli sulla strategia dell'Intelligenza artificiale"*, con questa nuova declinazione: *"sistemi software (ed eventualmente hardware)*

---

[*] Il presente scritto è stato elaborato nel contesto di una ricerca Astrid su Intelligenza artificiale e diritto, ed è stato pubblicato in ASTRID, "Intelligenza artificiale e diritto: una rivoluzione? Amministrazione, responsabilità, giurisdizione", a cura di Filippo Donati, Alessandro Pajno, Antonio Perrucci, vol. II, Ed. il Mulino, Bologna, 2022

[**] *Avvocato / comitato scientifico Fondazione Leonardo - Civiltà delle Macchine / professore a contratto di Diritto penale dell'ambiente, del lavoro e della sicurezza informatica presso l'Università Ca' Foscari di Venezia*

[1] G. Ubertis, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, in *Diritto penale contemporaneo*, 2020, 4, 76.

[2] https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:52018DC0237&from=IT.

progettati dall'uomo che, dato un obiettivo complesso, agiscono nella dimensione fisica o digitale percependo il proprio ambiente attraverso l'acquisizione di dati, interpretando i dati strutturati o non strutturati raccolti, ragionando sulle conoscenze, o elaborando le informazioni derivate da questi dati e decidendo le migliori azioni da intraprendere per raggiugere l'obiettivo dato. I sistemi d'IA possono usare regole simboliche o apprendere un modello numerico, e possono anche adattare il loro comportamento analizzando come l'ambiente è influenzato dalle loro azioni precedenti"[3].

Lo scorso 21 aprile la Commissione Europea ha pubblicato la propria *Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione*[4] e relativi Allegati[5].

Nella Relazione introduttiva e nel terzo *considerando* di tale proposta di Regolamento, l'intelligenza artificiale viene descritta come *"una famiglia di tecnologie in rapida evoluzione in grado di apportare una vasta gamma di benefici economici e sociali in tutto lo spettro delle attività industriali e sociali"*. Il successivo sesto *considerando* ricorda che *"la definizione dovrebbe essere basata sulle principali caratteristiche funzionali del software, in particolare sulla capacità, per una determinata serie di obiettivi definiti dall'uomo, di generare output quali contenuti, previsioni, raccomandazioni o decisioni che influenzano l'ambiente con cui il sistema interagisce, tanto in una dimensione fisica quanto in una dimensione digitale"*.

La definizione "autentica" d'intelligenza artificiale è cristallizzata all'art. 3, n. 1, della proposta di Regolamento, nei termini di *"un software sviluppato con una o più delle tecniche e degli approcci elencati nell'allegato I, che può, per una determinata serie di obiettivi definiti dall'uomo, generare output quali contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono"*.

Il richiamo all'allegato I, poi, descrive in termini analitici i succitati approcci, come facenti riferimento all'apprendimento automatico, supervisionato e non, basati sulla logica, sulla conoscenza, sulla programmazione induttiva e sulle inferenze deduttive o statistiche.

---

[3] https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_it.pdf.

[4] https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0006.02/DOC_1&format=PDF

[5] https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0006.02/DOC_2&format=PDF

Acquisizione e interpretazione di dati, elaborazione d'informazioni, decisioni d'azione rispetto ad uno scopo e successivo adattamento evolutivo, sulla scorta delle assunzioni decisorie precedenti: non vi è dubbio che si tratta di uno schema definitorio perfettamente in grado di contenere anche il processo conoscitivo tipico della decisione giudiziaria.

Non a caso, la Commissione Europea per l'Efficacia della Giustizia (CEPEJ) del Consiglio d'Europa, con l'emanazione della "*Carta etica europea per l'uso dell'intelligenza artificiale nei sistemi di giustizia penale e nei relativi ambienti"*, del 3 dicembre 2018, aveva già fornito una definizione d'intelligenza artificiale specificamente declinata su tale materia, nei termini dell' *"insieme di metodi scientifici, teorie e tecniche finalizzate a riprodurre mediante le macchine le capacità cognitive degli esseri umani. Gli attuali sviluppi mirano a far svolgere alle macchine compiti complessi precedentemente svolti da esseri umani"*[6].

La capacità – meglio la possibilità, visto che la capacità da un punto di vista tecnico sembra conclamata – di trasferire in capo ad una macchina l'elaborazione di un processo conoscitivo (e una conseguente decisione), che è poi destinato a produrre ricadute pregiudizievoli sui beni primari dell'essere umano, quali la sua libertà, riservatezza, non discriminazione e accesso paritario ai diritti e alle facoltà della comunità politica di appartenenza: questo è il quesito che impegna, inevitabilmente con grande difficoltà, l'etica, il diritto e la politica.

Da un lato, infatti, non vi è dubbio che non avrebbe alcun senso, risulterebbe antistorico e razionalmente di retroguardia, pensare che l'universo del diritto penale possa o debba rimanere estraneo alla penetrazione dell'intelligenza artificiale. L'approccio della politica dell'Unione Europea sul punto è chiaramente indicato nel già citato recente *Libro bianco*: la tecnologia digitale è parte centrale di tutti gli aspetti della vita dei cittadini europei e l'atteggiamento rispetto ai contributi della stessa non può non essere di fiducia, nella consapevolezza della sua indubbia capacità di produrre vantaggi di carattere economico, sociale, assistenziale e d'interesse pubblico.

Il sistema giudiziario, come già ricordato, si presta perfettamente a rendere operativi gli algoritmi complessi dei sistemi informatici di apprendimento[7]; in fin dei conti, l'universo dell'applicazione del diritto vive di esperienze conoscitive e

[6] https://rm.coe.int/carta-etica-europea-sull-utilizzo-dell-intelligenza-artificiale-nei-si/1680993348.

[7] S. Riondato, *Robot: talune implicazioni di diritto penale*, in P. Moro - C. Sarra (a cura di), *Tecnodiritto. Temi e problemi d'informatica e robotica giuridica*, Milano, 2017, p. 85 ss.

predittive assolutamente sovrapponibili a quelle con le quali operano le cosiddette macchine intelligenti.

Dall'altro lato, tuttavia, la giurisdizione penale è chiamata a rispondere a tutta una serie di requisiti standard, anche di matrice costituzionale, che risultano coinvolgere in maniera indisponibile il dominio umano del processo decisorio, che, non lo si dimentichi, finisce sempre per intaccare i diritti primari dei suoi destinatari.

La soggezione del giudice *"soltanto alle legge"*, il contraddittorio paritario nella formazione delle prove, l'obbligo di motivazione e di giustificazione razionale della sentenza, il diritto d'impugnazione avverso la stessa[8], rischierebbero infatti di perdere di significato precettivo, al cospetto di una decisione automatizzata, in quanto tale non assunta liberamente, non giustificabile e pertanto potenzialmente irriformabile.

Inoltre, il giudice del processo penale è sempre libero nel suo convincimento, che può formarsi anche a proposito di prove di natura meramente indiziaria, il cui apprezzamento qualitativo – la loro gravità, precisione e concordanza – si compie anche attraverso intuizioni, emozioni, percezioni di sfumature probatorie, che sembrano essere tipicamente umane[9] e non accessibili ad un sistema automatico, ancorché intelligente.

Alla stessa stregua, nel giudizio penale la regina delle prove è quella testimoniale. Per quanto possa essere sterminata la quantità d'informazioni che il sistema automatico intelligente riesce a processare, la valutazione sulla credibilità di una testimonianza, soprattutto da un punto di vista soggettivo, è esperienza che sembra troppo umana per poter essere colta da una macchina.

In termini teorici, inoltre, le difficoltà a coniugare in modo fecondo intelligenza artificiale e giudizio di responsabilità penale, attengono alla questione delle fondamenta etiche della materia in esame, che impongono degli apprezzamenti di carattere morale – sembrerebbe – incompatibili con la fisiologica e strutturale a-moralità della macchina artificiale.

Ci si confronti, infine, con il tema del "dubbio" sulla responsabilità dell'imputato; esso impone al giudice penale di assolvere, perché così stabilisce il codice di procedura penale. Il modello liberal-costituzionale del nostro ordinamento, infatti,

[8] G. Ubertis, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, cit., 83.

[9] V. Manes, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in *dis CRIMEN*, 2020, pp. 1-22, http://www.antoniocasella.eu/archica/Manes_oracolo.algoritm_15mag20 .pdf.

impone in capo alla pubblica accusa una prova piena e senza riserve della responsabilità dell'accusato.
Il sistema artificiale, a differenza dell'uomo, è rapido, efficiente, ed offre certezze; non si sofferma e non perde tempo, a dubitare delle proprie conclusioni: difficile pensare che esso possa essere compatibile con un sistema che fa del dubbio il proprio fondamento conoscitivo.
Quando i dubbi sono superati, poi, se lo sono nella direzione dell'affermazione di responsabilità dell'imputato, la condanna ad una determinata sanzione, contiene in sé, tra le tante funzioni che la pena esprime, un rimprovero etico corrispondente all'apprezzamento specifico di un determinato grado di colpevolezza[10]. Tutte le funzioni della pena, a ben vedere, rispecchiano esigenze di carattere etico.
È complicato pensare che la macchina, anche se intelligente, possa cogliere nel dettaglio la dimensione del rimprovero soggettivo di una determinata scelta sanzionatoria, che è invece centrale nella quantificazione della pena.
È dunque all'interno di questo antagonismo, tra crescita inarrestabile degli spazi applicativi dell'intelligenza artificiale, *versus* umanità – almeno apparentemente – irrinunziabile della giustizia penale, che può essere sviluppato un percorso d'analisi degli ambiti di penetrazione in essa dell'intelligenza artificiale, per analizzarne potenzialità, limiti, ricadute rischiose ed esigenze di protezione rispetto a queste ultime.
L'indicazione propositiva del diritto europeo, da questo punto di vista, è già chiara. Il quarantesimo *considerando* della proposta di Regolamento UE sulla disciplina dell'intelligenza artificiale indica espressamente come *"alcuni sistemi di IA destinati all'amministrazione della giustizia e ai processi democratici dovrebbero essere classificati come sistemi ad alto rischio, in considerazione del loro impatto potenzialmente significativo sulla democrazia, sullo Stato di diritto, sulle libertà individuali e sul diritto a un ricorso effettivo e a un giudice imparziale"*. Conseguentemente, la predetta classificazione ad alto rischio deve necessariamente coinvolgere tutti *"i sistemi di IA destinati ad assistere le autorità giudiziarie nelle attività di ricerca e interpretazione dei fatti e del diritto e nell'applicazione della legge a una serie concreta di fatti"*[11].

[10] A. Pagliaro, *Responsabilità oggi (nessi e interdipendenze tra responsabilità politica, responsabilità giuridica e responsabilità morale)*, in A. Pahliaro, *Il diritto penale tra norma e società - Scritti 1956-2008*, IV, 2009, p. 615.
[11] Infatti, l'Allegato III alla proposta di Regolamento classifica espressamente ad alto rischio, al punto 8, lett. a), proprio *"i sistemi di IA destinati ad assistere un'autorità giudiziaria nella ricerca e nell'interpretazione dei fatti e del diritto e nell'applicazione della legge a una serie concreta di fatti"*. Ad analoga stregua, identica classificazione coinvolge (punto 6, lett. a) *"i*

2. *Lo spazio (già occupato) dei sistemi d'intelligenza artificiale nel procedimento penale*

Molteplici sono gli ambiti nei quali, a proposito dell'amministrazione della giustizia, possono trovare applicazione le nuove frontiere dell'intelligenza artificiale.
Possiamo, per praticità di trattazione della materia, suddividere i contesti di potenziale sviluppo dei sistemi d'intelligenza artificiale in ambito penalistico in due macro-categorie, che sono l'analisi predittiva rispetto al potenziale esito di un giudizio, nonché la formulazione di un giudizio e la determinazione di contenuto di una decisione giudiziaria[12].

2.1. *Analisi predittiva dell'esito di un giudizio*

Sul piano processuale, la capacità algoritmica "intelligente" di gestione di dati, che può produrre sintesi comparatistiche di tutti i precedenti giudiziali per ciascuna singola fattispecie di reato, in fatto e in diritto, può manifestarsi nella capacità di prevedere l'esito di un giudizio.
Il portale francese *Predictice*[13], che nasce a Parigi come start-up nel 2016, offre commercialmente il proprio prodotto agli operatori giuridici, indicando la sua capacità di aggiornamento in tempo reale e di processo informatico di *"milioni di decisioni in un secondo"*; esso si propone di formulare previsioni sull'esito di un

sistemi di IA destinati a essere utilizzati dalle autorità di contrasto per effettuare valutazioni individuali dei rischi delle persone fisiche al fine di determinare il rischio di reato o recidiva in relazione a una persona fisica o il rischio per vittime potenziali di reati", (lett. d) *"i sistemi di IA destinati a essere utilizzati dalle autorità di contrasto per la valutazione dell'affidabilità degli elementi probatori nel corso delle indagini o del perseguimento di reati"*, (lett. f) *"i sistemi di IA destinati a essere utilizzati dalle autorità di contrasto per la profilazione delle persone fisiche di cui all'articolo 3, paragrafo 4, della direttiva (UE) 2016/680 nel corso dell'indagine, dell'accertamento e del perseguimento di reati"* e (lett. g) *"i sistemi di IA destinati a essere utilizzati per l'analisi criminale riguardo alle persone fisiche, che consentono alle autorità di contrasto di eseguire ricerche in set di dati complessi, correlati e non correlati, resi disponibili da fonti di dati diverse o in formati diversi, al fine di individuare modelli sconosciuti o scoprire relazioni nascoste nei dati"*.

[12] Oltre all'attività di analisi predittiva di determinati eventi patologici a fini preventivi, che non è oggetto di trattazione in questo contributo.

[13] https://predictice.com/.

futuro giudizio, tenendo conto, nell'ambito dell'analisi degli esiti di tutti i precedenti rilevanti, di una molteplicità molto significativa di parametri, tra i quali la giurisdizione attivata, il singolo magistrato coinvolto, gli studi legali che hanno patrocinato le parti.

Si tratta di una potenzialità effettivamente molto interessante per gli operatori giuridici, pubblici e privati.

Poter conoscere, sulla base di un processo algoritmico, tutta una serie d'informazioni sul comportamento del giudice nelle precedenti vicende analoghe a quelle d'interesse, ad esempio, può certamente orientare con intelligenza supplementare le parti, nella scelta della strategia processuale da adottare.

Peraltro, essi possono rivelarsi particolarmente funzionali a migliorare la stessa attività del giudice, che è fortemente agevolato nel conoscere a fondo gli orientamenti della giurisprudenza e nell'analizzarne le ragioni di dettaglio. Tale livello di approfondimento può senza dubbio contribuire a favorire la coerenza – e quindi la conoscibilità da parte dei cittadini – dello svolgersi della giurisdizione. Il giudice, con l'ausilio di tali strumenti, può infatti determinarsi in modo maggiormente informato nell'assumere le proprie decisioni, confrontandosi in modo aperto con gli orientamenti dominanti e quelli minoritari, in funzione dell'accertamento di quelli più adeguati alla soluzione del caso concreto.

Nello spazio giuridico europeo, ove oramai la legalità penale è legata all'individuazione della prevedibilità e dell'accessibilità del comando – anche di origine giurisprudenziale – da parte del destinatario di esso, anche *a latere* o oltre la legalità legislativa formale, è indubbio che l'incremento della conoscenza accessibile per il mezzo dell'intelligenza artificiale può contribuire a rafforzare la stabilità dell'intero ordinamento.

2.2. *Formulazione del giudizio e determinazione del contenuto di una sentenza*

In questo contesto di sviluppo dirompente delle potenzialità applicative del *machine learning* e dell'intelligenza artificiale, è sostanzialmente inevitabile che la sua penetrazione nel mondo del diritto finisca per attingere anche quell'attività conoscitiva normalmente considerata solo umana, che è quella della formulazione

dei giudizi e della determinazione di contenuto concreto di una specifica decisione giudiziaria.

Nonostante le evidenti peculiarità della giurisdizione penale (già ricordate in apertura di questo contributo), che rendono non immediatamente riconoscibile uno spazio applicativo per le decisioni automatizzate, a ben vedere sono molte le realtà nelle quali già viene fatto ampio uso dell'intelligenza artificiale e degli algoritmi predittivi, con particolare riferimento allo svolgimento delle prognosi di pericolosità sociale, alla determinazione del contenuto delle misure di sicurezza e di prevenzione e finanche alla commisurazione della pena in fase di cd. *sentencing*.

Negli Stati Uniti d'America, infatti, già da una decina d'anni risultano operativi algoritmi predittivi della pericolosità criminale, che sviluppano il processo di *Violence Risk Assessement* a partire dall'elaborazione statistica delle informazioni raccolte a proposito di specifici e predeterminati fattori di rischio: dall'età, al sesso, all'etnia di appartenenza, alla residenza, al contesto familiare, alla posizione economico-lavorativa, al consumo di alcolici o stupefacenti.

Il risultato dell'elaborazione automatizzata determina l'attribuzione di un punteggio finale al soggetto analizzato, che fonda una pluralità di decisioni giudiziali, da quella sulla liberazione o meno dell'indiziato su cauzione, alla sua ammissione alla messa alla prova, all'accesso alle misure alternative alla detenzione.

Particolarmente diffuso, in ambito cautelare, è il sistema d'intelligenza artificiale denominato *Public Safety Assessment – PSA*[14], già in uso in una pluralità di giurisdizioni degli Stati Uniti (*Arizona*, *Kentucky*, *New Jersey*, in città molto importanti come *Phoenix*, *Chicago* e *Houston*), che gestisce un database di oltre un milione e mezzo di casi precedenti, riferiti ad otre trecento diverse giurisdizioni in tutto il territorio U.S.A.

Tale algoritmo, incrociando i dati disponibili con nove fattori di misurazione dei rischi, quali l'età, il titolo di reato dell'arresto, i precedenti penali e di polizia (mentre non sono contemplati indicatori sull'etnia di appartenenza e sulla provenienza geografica), attribuisce un punteggio finale, sulla scorta del quale il giudice decide a proposito dell'applicazione dell'istituto del *bail*, vale a dire la liberazione o meno dell'arrestato su cauzione.

Il sistema genera un *report*, che include i punteggi e le proprie valutazioni quanto ai fattori di rischio, suggerendo le proprie condizioni di rilascio preliminare. Esso viene fornito al giudice, e, nella maggior parte delle giurisdizioni, è messo a disposizione del procuratore e dell'avvocato difensore.

[14] https://advancingpretrial.org/psa/about/.

Tuttavia, tale sistema artificiale non fornisce mai un risultato decisorio "finale", essendo sempre un giudice umano quello che si assume la responsabilità di farne uso e di decidere in che misura.

In ambito esecutivo, invece, *ampiamente* applicato è il software *Correctional Offender Management Profiling for Alternative Sanctions – COMPAS*, vale a dire un software della società privata *Equivant*, utilizzato, sin dal 1998, per la soluzione di oltre un milione di cause penali. La presentazione del prodotto, da parte della società titolare, è particolarmente significativa: *"the right data, in the right hands, at the right time. In the justice system, this can mean the difference between liberty and detention, efficiency and delay"*[15].

COMPAS si prefigge il compito di prevedere il rischio di recidiva di un determinato imputato e funziona attraverso un algoritmo che elabora le risposte a ben 137 quesiti, che fanno riferimento ad una pluralità molto articolata di questioni che lo riguardano: dai precedenti penali alle condizioni socio-economiche, dal livello d'istruzione alla stabilità residenziale, dal livello delle relazioni sociali alla dedizione alle sostanze alcoliche e stupefacenti[16]. Il produttore del software assicura che tra i dati elaborati automaticamente non possono rientrare in alcun modo temi di carattere etnico e razziale.

Sulla scorta di questo sistema intelligente, di *COMPAS* si è fatto uso anche in sede di *sentencing*, per determinare la commisurazione della pena a carico di un imputato riconosciuto colpevole di un determinato reato.

Nell'acceso dibattito circa l'opportunità o meno di fare uso di tale strumento decisorio, risulta interessante ricordare come del sistema *COMPAS* si sia occupata la Corte Suprema del Wisconsin, nel 2016, nel procedimento penale *State of Wisconsin /vs Loomis*.

Eric L. Loomis era stato condannato dalla Corte territoriale di *La Crosse* a sei anni di reclusione, per una sequenza di vari reati, riconnessi al fatto di essere stato colto alla guida di un'autovettura precedentemente usata durante una sparatoria con la Polizia.

Riconosciuta la responsabilità per i reati contestati, la Corte, ai fini del calcolo della pena, aveva ordinato un *Presentence Investigation Report – PSI*, vale a dire una relazione dei risultati delle investigazioni condotte sulla storia personale dell'imputato, finalizzata a verificare la presenza di circostanze significative per la determinazione complessiva della pena stessa. Il *PSI* acquisito dalla Corte si

[15] https://www.equivant.com/.

[16] T. Brennan - W. Dieterich - B. Ehret, *Evaluating the Predictive Validity of the Compas Risk and Needs Assessment System*, in *Criminal Justice and Behavior*, 36, 2009, pp. 22-23.

avvaleva anche dei risultati elaborati da *COMPAS*, che avevano dato di *Loomis* un giudizio di soggetto ad alto rischio di recidiva.
Sulla scorta di tale risultato il Tribunale territoriale decideva di fare uso degli esiti conoscitivi di *COMPAS*, sia per determinare l'ammontare della pena finale, sia per negare all'imputato la libertà vigilata.
La vicenda, per iniziativa del giudice d'appello veniva rimessa all'attenzione della Corte Suprema del *Wisconsin*, proprio per decidere sulla fondatezza della censura prospettata dall'imputato, che aveva stigmatizzato l'utilizzo dell'algoritmo *COMPAS* nel giudizio che lo aveva riguardato.
I profili di doglianza dell'imputato erano vari e articolati. L'algoritmo era stato utilizzato per la quantificazione della pena, quando si trattava di un software finalizzato alle sole prognosi di recidiva. La sanzione applicata, in quanto determinata da una macchina, non dava garanzie di essere effettivamente e correttamente individualizzante. Risultava violato il diritto al giusto processo, perché la condanna non poteva fondarsi su informazioni accurate, visto che il funzionamento del sistema informatico è coperto dal segreto, a garanzia dei diritti di proprietà industriale sullo stesso. L'algoritmo, infine, mostrava il limite del pregiudizio di genere e di carattere razziale.
La Corte Suprema del *Wisconsin* rigetta il ricorso di *Eric L. Loomis*[17], evidenziando come, nonostante la segretezza del software, il manuale sul funzionamento di *COMPAS* sia pubblico e trasparente nell'indicazione dei criteri di attribuzione dei punteggi. Quanto all'accuratezza nella raccolta dei dati da parte del sistema, preesistenti studi autorevoli avevano concluso nel senso che – pur non perfetto – *COMPAS* rappresentava un mezzo di calcolo comunque affidabile.
Da ultimo, quanto al tema essenziale dell'equità del processo, la Corte Suprema evidenzia che lo stesso in realtà non si pone: infatti, le valutazioni di *COMPAS* non erano state esclusive nella determinazione del contenuto della decisione, ma certamente sottoposte al controllo e alla validazione di un giudice umano, che le aveva fatte proprie.
Peraltro, nonostante il rigetto del ricorso dell'imputato, la Corte del *Wisconsin* non ha inteso avallare senza compromessi la legittimità dell'uso giudiziale di *COMPAS* nell'attività di *sentencing*. Infatti, la stessa, con un vero e proprio *warning*, ha messo in evidenza i limiti e le cautele che sono necessarie affinché risulti legittimo, per il giudice umano, l'utilizzo del sistema algoritmico.

[17] STATE of Wisconsin, Plaintiff-Respondent vs Eric L. Loomis, Defendant-Appellant - Supreme Court of Wisconsin. Argued April 5, 2016. Decided July 13, 2016, in https://www.courts.ca.gov/documents/BTB24-2L-3.pdf.

I principali elementi negativi che la Corte evidenzia attengono alla segretezza del sistema quanto al suo metodo di funzionamento e al fatto che il medesimo effettua valutazioni aggregate su base collettiva e non di tipo individuale, con conseguente rischio di sovrastima della recidiva, in relazione a specifiche minoranze etniche.
Ne consegue che è dovere del giudice operare la propria autonoma valutazione discrezionale dei risultati offerti da *COMPAS*, bilanciando gli stessi con altri fattori di apprezzamento tipicamente umano. In questa prospettiva, molto chiara risulta l'opinione concorrente del giudice Drake Roggensack: *"I write to clarify that while our holding today permits a sentencing court to consider COMPAS, we do not conclude that a sentencing court may rely on COMPAS for the sentence it imposes. Because at times the majority opinion interchangeably employs consider and rely when discussing a sentencing court's obligations and the COMPAS risk assessment tool, our decision could mistakenly be read as permitting reliance on COMPAS"*.

3. *La politica e il diritto dell'Unione Europea sullo spazio dell'intelligenza artificiale nell'azione giudiziaria e di polizia in ambito penale*

La breve rassegna che è stata svolta evidenzia la massiccia penetrazione dell'intelligenza artificiale nel mondo della prevenzione di polizia e dell'amministrazione della giustizia penale.
Essa è destinata certamente ad incrementarsi, di pari passo con lo sviluppo inarrestabile del progresso tecnologico ed informatico.
Tale tendenza non può non essere interpretata positivamente, in una logica fiduciaria nei confronti degli innegabili effetti validi della crescita dell'automazione, che può certamente migliorare l'efficienza del sistema giudiziario: soprattutto sul versante della diminuzione del tempo dei processi decisori e del contenimento dei costi, in una prospettiva di miglior capacità di allocazione delle risorse disponibili.
Le questioni problematiche che emergono sul tappeto, tuttavia, sono molteplici, e necessitano della messa in campo di tutta una serie di strumenti di salvaguardia e tutela, in grado di scongiurare quei plurimi effetti potenzialmente negativi, già emersi in controluce nella disamina sin qui svolta, che si mostrano incompatibili con il sistema dei diritti e delle garanzie approntati dalle strutture costituzionali degli ordinamenti nazionali e soprattutto dalla politica e dal diritto del processo d'integrazione europea.
Le istituzioni europee, da questo punto di vista, si mostrano come le più feconde, non solo nell'evidenziare l'irreversibilità dell'approccio positivo, in termini di

eccellenza e fiducia, alla crescita di spazio applicativo dell'intelligenza artificiale – si pensi al già citato Libro Bianco del 2020 –, ma anche nel sottolineare i "segnali d'allarme" che impongono cautele e attenzioni, rispetto al rischio concreto di degenerazioni applicative, potenzialmente pregiudizievoli per la salvaguardia dei principi fondamentali del diritto dell'Unione.

Già nel 2017 il Parlamento Europeo, con il proprio *Report on fundamental rights implications of big data: privacy, data protection, non discrimination, security and law-enforcement*, seppur con particolare riferimento alle attività di polizia predittiva, aveva posto la questione problematica del rischio di pregiudizio per i diritti di non discriminazione[18].

La questione è stata poi ripresa dalla già citata "*Carta etica europea per l'uso dell'intelligenza artificiale nei sistemi di giustizia penale e nei relativi ambienti"* del 2018, che ha inteso declinare, a fronte della *"crescente importanza della intelligenza artificiale (IA) nelle nostre moderne società e dei benefici attesi quando questa sarà pienamente utilizzata al servizio della efficienza e qualità della giustizia"*, le linee guida e i principi irrinunciabili nell'applicazione dell'intelligenza artificiale al diritto, cui sono chiamati ad attenersi tutti *"i soggetti pubblici e privati responsabili del progetto e sviluppo degli strumenti e dei servizi della IA"*, individuandoli esplicitamente nel rispetto dei diritti fondamentali, di non discriminazione, di qualità e sicurezza, di trasparenza, e da ultimo, ma certamente di primaria importanza, del principio di garanzia dell'intervento umano.

Quest'ultima indicazione, quale applicazione del principio dell'IA *"under user control"*, viene ricordata come specificamente finalizzata a *"precludere un approccio deterministico"* (*preclude a prescriptive approach*) e ad *"assicurare che gli utilizzatori agiscano come soggetti informati ed esercitino il controllo delle scelte effettuate" (ensure that users are informed actors and in control of the choices made*).

L'enunciato, pur nella sua essenzialità, nel riconoscere la più ampia possibilità di utilizzo dell'intelligenza artificiale nell'ambito della giustizia penale, detta tuttavia due precise condizioni pregiudiziali, vale a dire che gli operatori siano soggetti qualificati all'uso dei sistemi algoritmici e che ogni decisione sia sottoposta al controllo umano (ad esempio, da parte del giudice utilizzatore del sistema automatizzato).

---

[18] https://www.europarl.europa.eu/doceo/document/A-8-2017-0044_EN.html, ove è espressamente indicato come "*low-quality data and/or low-quality procedures behind decision-making processes and analytical tools could result in biased algorithms, spurious correlations, errors, an underestimation of the legal, social and ethical implications"*.

Solo in questi termini può essere prevenuto il già citato rischio di *"approccio deterministico"*, che altro non è che un eccessivo automatismo e standardizzazione delle decisioni, che potrebbero finire per sottrarsi all'apprezzamento delle specificità di ciascun caso concreto.

Tale indicazione prescrittiva, tuttavia, non si mostra come una novità nel panorama del vigente diritto europeo e nazionale.

Già l'art. 22 del Reg. UE 2016-679, *General Data Protection Regulation – GDPR*[19], così come completato dalle Linee Guida *Art. 29 Data Protection Working Party* n. 251 del 3 ottobre 2017[20] sul processo decisionale automatizzato relativo alle persone fisiche e sulla loro profilazione, elaborate dal Comitato Europeo per la protezione dei dati personali, aveva fissato il diritto di ogni soggetto interessato a non essere sottoposto a procedimento e a non dover subire gli effetti giuridici di decisioni basate unicamente sul trattamento automatizzato, essendo sempre necessario, per converso (salve talune eccezioni), un intervento umano, nel processo istruttorio, valutativo e decisionale, effettivo e non meramente formale.

Quanto, infine, alla già ricordata proposta di Regolamento UE dello scorso aprile, in forza della classificazione "ad alto rischio" dei sistemi d'intelligenza artificiale utilizzabili nell'ambito della decisione penale, per gli stessi è stato ipotizzato un dovere di progettazione e sviluppo (art. 13) nel rispetto della prescrizione della *"trasparenza"*, nel senso che i medesimi devono *"garantire che il loro funzionamento sia sufficientemente trasparente da consentire agli utenti di interpretare l'output del sistema e utilizzarlo adeguatamente"*, sia di quella della *"sorveglianza umana"*, intesa come comprensiva di *"strumenti di interfaccia uomo-macchina adeguati, in modo tale da poter essere efficacemente supervisionati da persone fisiche durante il periodo in cui il sistema di IA è in uso"*.

In ambito nazionale, poi, è vigente il d.lgs. 18 maggio 2018, n. 51[21], recante Attuazione della direttiva (UE) 2016/680 del Parlamento europeo e del Consiglio, del 27 aprile 2016, 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, nonché alla libera circolazione di tali dati.

L'art. 8 del d.lgs. in esame, a ben vedere, riproducendo sostanzialmente l'art. 22 del *GDPR*, prescrive espressamente il divieto di decisioni interamente

[19] https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32016R0679&from=EN.
[20] https://ec.europa.eu/newsroom/article29/items/612053.
[21] https://www.gazzettaufficiale.it/eli/id/2018/05/24/18G00080/sg.

automatizzate, ma con l'indicazione che tale divieto opera solo se esse producono effetti negativi nei confronti dell'interessato. Ne deriva che, in campo penale, le decisioni totalmente automatizzate appaiono potenzialmente legittime se producono effetti favorevoli (o comunque che vengono considerati tali) per il loro destinatario.
Poi, sempre in analogia con le indicazioni del *GDPR*, pur essendo richiamata la possibilità della decisione automatizzata nei casi espressamente previsti dalla legge, è comunque fatto salvo il diritto dell'interessato a richiedere l'intervento umano da parte del titolare del trattamento.

4. *La risoluzione del Parlamento europeo del 6 ottobre 2021 "sull'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale":* warning *o grido d'allarme?*

Con l'adozione di tale atto di indirizzo[22], il Parlamento europeo ha assunto decisa e risoluta posizione, formalmente trasmessa al Consiglio e alla Commissione, sui significativi fattori di concreto pregiudizio, per la tutela dei diritti fondamentali nello spazio giuridico europeo, che può comportare l'utilizzo dell'IA in materia penale di polizia.
Le *"promesse"* assicurate dall'IA, che *"può potenzialmente diventare parte integrante del nostro ecosistema di giustizia penale fornendo analisi investigative e assistenza"*, sono straordinarie; ma – esordisce nel suo primo *considerando* il Parlamento europeo – esse recano con sé altrettanti *"rischi enormi per i diritti fondamentali e le democrazie basate sullo Stato di diritto"*.
Il Parlamento volge il suo sguardo oltre il territorio del continente europeo, e rileva come in diversi paesi l'IA è utilizzata sia in molteplici attività di polizia di contrasto al crimine (dalle tecnologie di riconoscimento facciale, all'identificazione vocale, alle tecnologie di lettura labiale, all'analisi dei segnali acustici), sia in ambito giudiziario (decisioni in materia di custodia cautelare, irrogazione delle pene, quantificazione del rischio di recidiva, gestione dei precedenti giurisprudenziali).
Osserva però che, quanto all'attività di polizia, molti di questi strumenti *"sarebbero illegali ai sensi dell'acquis dell'Unione in materia di protezione dei*

[22] https://www.europarl.europa.eu/doceo/document/TA-9-2021-0405_IT.html.

dati e della relativa giurisprudenza". Ad analoga stregua, in relazione agli accertamenti giudiziari, l'utilizzo dell'IA, oltre a causare distorsioni e ridurre le opportunità per i soggetti appartenenti a minoranze, *"comporta una serie di rischi potenzialmente elevati, e in alcuni casi inaccettabili, per la protezione dei diritti fondamentali degli individui"*, *"della vita privata e dei dati personali, ... della libertà di espressione e informazione, la presunzione di innocenza, il diritto a un ricorso efficace e a un processo equo nonché rischi per la libertà e la sicurezza degli individui"*.

Rispetto a taluni approcci applicativi stranieri, dunque, l'indicazione del Parlamento europeo è tranciante: essi interferiscono in modo sproporzionato con i diritti fondamentali e non possono essere seguiti dall'U.E.

Specifica attenzione sul punto è poi spesa a proposito dei meccanismi d'identificazione basati su algoritmi, che, inciampando ancora in un numero significativo di errori di classificazione, offrono risultati distorti e discriminatori. Per converso, per il diritto dell'Unione *"gli individui non hanno soltanto il diritto a essere identificati correttamente, ma anche di non essere identificati, salvo quando richiesto per legge per interessi pubblici imperativi e legittimi"*.

Osserva inoltre l'Assemblea di Strasburgo che i sistemi d'IA utili ati in attività di polizia e di accertamento giudiziario sono anch'essi vulnerabili, sia volontariamente attraverso la pratica del cd. *"avvelenamento dei dati"*, sia anche solo per semplice inefficienza e capacità di protezione, rispetto a possibili attacchi informatici, che possono determinare fughe di dati o comunque violazioni della loro sicurezza attraverso accessi non autorizzati. E siccome il risultato fornito dagli applicativi in esame *"è necessariamente influenzato dalla qualità dei dati utilizzati"*, le distorsioni discriminatorie *"possono essere intrinseche ai sistemi di dati di base"* e amplificarsi senza controllo, attraverso il perpetuarsi del loro utilizzo.

Da ultimo, con specifico riferimento a taluni applicativi d'IA in uso o in fase d'implementazione anche in territorio U.E., il Parlamento segnala alle istituzioni europee la propria *"profonda preoccupazione"*.

Ricorda infatti che la banca dati privata *ClearviewAI*, che custodisce oltre tre miliardi di immagini di individui – anche cittadini europei – illegalmente raccolte dai social network e da altre fonti internet, viene utilizzata in talune realtà per il riconoscimento delle persone; e che tale degenerazione meriterebbe la prescrizione di un esplicito divieto europeo di utilizzo di database privati nelle attività di polizia.

Quanto, poi, al progetto *iBorderCTRL – Intelligent portable border control system*, che vorrebbe implementare un sistema di analisi dei movimenti facciali per il

controllo delle frontiere, pur finanziato dalla Commissione europea nell'ambito del progetto *Horizon 2020* e oggetto di sperimentazione in alcuni paesi dell'Unione (Ungheria, Lettonia e Grecia), il Parlamento ne rileva l'incompatibilità strutturale con il diritto U.E. e invita la Commissione stessa sia ad interrompere il finanziamento della ricerca in corso, sia a determinarsi a vietare ogni forma di trattamento di dati biometrici (comprese le immagini facciali), che possa condurre a meccanismi di sorveglianza di massa negli spazi aperti al pubblico.

I segnali d'allarme che il Parlamento europeo evidenzia, dunque, sono molti, significativi e decisamente grevi. Rispetto ad essi, pertanto, viene posta con chiarezza e trasparenza la questione centrale, che attiene alla corretta declinazione del rapporto tra possibilità offerte dalla tecnica e indicazioni della politica: queste ultime devono sempre essere poste in via pregiudiziale e le prime devono essere utilizzate se e in quanto serventi alla realizzazione di un determinato obiettivo politico prefissato. Come si legge testualmente nel diciassettesimo *considerando*, *"la diffusione dell'IA nel settore delle attività di contrasto e nel settore giudiziario non dovrebbe essere considerata una mera questione di realizzabilità tecnica ma piuttosto una decisione politica riguardante la progettazione e gli obiettivi dei sistemi di attività di contrasto e di giustizia penale"*. Infatti, *"il moderno diritto penale si basa sull'idea che le autorità reagiscono ad un reato dopo che è stato commesso, senza supporre che le persone siano pericolose e debbano essere sorvegliate costantemente per prevenire possibili illeciti"*.

I termini della questione sono posti in modo molto limpido: lo spazio per gli strumenti dell'IA in ambito penale e di polizia, per essere effettivamente funzionale, infondere semplificazione, corretta gestione delle risorse e fiducia nei destinatari di esso, dev'essere predeterminato dagli obiettivi della politica, in coerenza con i limiti da essa imposti e in modo servente alla loro realizzazione.

In conseguenza di tale approccio, l'Assemblea parlamentare europea indica tutta una serie di pre-requisiti normativi che devono essere fissati per individuare l'ambito d'intervento dell'IA nell'universo della giurisdizione penale e dell'attività di polizia.

Il primo monito è di carattere generale: *"l'utilizzo dell'IA deve essere proibito se incompatibile con i diritti fondamentali"* e deve *"rispettare appieno i principi di dignità umana, non discriminazione, libertà di movimento, presunzione d'innocenza e diritto di difesa, compreso il diritto di non rispondere, libertà di espressione e informazione, libertà di riunione e associazione, uguaglianza dinanzi alla legge, principio dell'uguaglianza delle armi e diritto a un ricorso effettivo e a un processo equo"*.

L'essenzialità per il diritto U.E. della tutela della vita privata e della protezione dei dati personali, poi, determina che *"dovrebbe essere impedita la possibile identificazione degli individui da parte di un'applicazione di IA sulla base di dati precedentemente anonimizzati"*.
In questa cornice, i sistemi d'IA devono *"essere non discriminatori, sicuri e ... le relative decisioni devono essere spiegabili e trasparenti e rispettare l'autonomia umana e i diritti fondamentali, per poter essere considerati affidabili"*, secondo una logica funzionale al rispetto dei *"principi di necessità e proporzionalità"*.
Da questo punto di vista, *"l'applicazione generalizzata dell'IA ai fine della sorveglianza di massa sarebbe sproporzionata"* e l'uso delle applicazioni a ciò funzionali dovrebbe essere vietato.
Tuttavia, il tema sul quale il Parlamento europeo sviluppa le sue considerazioni più pregnanti è quello che attiene all'imprescindibilità del controllo umano sugli strumenti e i prodotti dell'intelligenza artificiale, che deve coinvolgere tutti i soggetti partecipanti della loro creazione e del loro utilizzo.
Anzitutto, *"le strutture aziendali che producono e gestiscono i sistemi d'IA"*, che devono *"assicurare trasparenza"* nel loro processo creativo, devono rispondere a requisiti di tracciabilità, spiegabilità e verifica, che sono incompatibili con la logica dei *"sistemi chiusi e contrassegnati come proprietari da parte dei fornitori"*.
Affianco alla *"sicurezza sin dalla progettazione"* degli strumenti algoritmici, viene altresì raccomandata la necessità di assicurare *"un controllo umano specifico prima di utilizzare determinate applicazioni"*.
A chi fosse coinvolto in un procedimento che prevede l'utilizzo, da parte dell'autorità, delle applicazioni di IA, dev'essere assicurato il diritto ad esserne informato, oltre che a poter *"accedere al processo di raccolta dei dati e quello relativo alle valutazioni correlate eseguite"*, quale estrinsecazione ineliminabile della garanzia di efficacia del diritto di difesa.
Da ultimo, *"se gli esseri umani fanno affidamento unicamente sui dati, i profili e le raccomandazioni generati dalle macchine, non saranno in grado di condurre una valutazione indipendente"*.
Infatti, ogni *"decisione che produce effetti giuridici o analoghi deve sempre essere presa da un essere umano, il quale possa essere ritenuto responsabile per le decisioni adottate"*, posto che *"ai sensi del diritto dell'UE, una persona ha il diritto di non essere sottoposta a una decisione che produca effetti giuridici o abbia effetti significativi nei suoi confronti fondata esclusivamente su un trattamento automatizzato di dati"*.

Peraltro, non essendo possibile eliminare il rischio del pregiudizio ai diritti fondamentali quale conseguenza dell'utilizzo dell'IA in ambito penale, fermo restando il principio di precauzione, che impone di considerare sempre che *"il primo e principale scopo deve innanzi tutto essere la prevenzione di tali conseguenze"*, è comunque necessario che venga definito *"un modello chiaro per attribuire la responsabilità per i potenziali effetti nocivi"* dell'IA, assicurando meccanismi per i quali *"la responsabilità giuridica e l'imputabilità devono sempre ricadere su una persona fisica o giuridica, che deve sempre essere identificata* – e qui la necessità del controllo umano sui risultati conoscitivi prodotti dai sistemi di IA è ribadita – *per le decisioni assunte con il sostegno dell'IA"*.

5. *Riflessioni conclusive*

Possiamo, a questo punto, trarre le fila di una prospettiva di sintesi, a proposito della disciplina giuridica dei rapporti tra intelligenza artificiale e diritto.
La diffusa e recente indicazione del Parlamento europeo in materia evidenzia la necessità di massima cautela nell'affidamento di processi decisori all'attività delle macchine, in uno con l'indicazione che gli strumenti automatizzati possono essere certamente utilizzati in sede giudiziaria, anche penale, ma solo in presenza di talune garanzie irrinunziabili, tra le quali spicca certamente l'obbligatorio intervento del controllo umano sul processo decisorio, sin dall'analisi dei meccanismi di funzionamento dell'algoritmo utilizzato.
Il quesito circa chi debba essere il soggetto istituzionale della disciplina in questo ambito, e quali siano le forme più consone, per una corretta regolamentazione di una materia così innovativa e delicata, trova per certi versi già risposta nelle scelte regolamentari, direttive e d'indirizzo europee che sono state analizzate.
Appare infatti del tutto incompatibile con un fenomeno globale, come quello dell'evoluzione tecnologica e informatica, il confinamento della disciplina di settore secondo specificazioni legislative di carattere nazionale.
Quanto al contenuto di tale regolamentazione normativa, la scelta di limitarla all'indicazione dei principi fondamentali, ai quali l'intelligenza artificiale deve conformarsi quando utilizzata in sede giudiziaria (rispetto dei diritti fondamentali, non discriminazione, qualità e sicurezza, trasparenza, imparzialità e correttezza, garanzia del controllo umano), sembra parimenti corretta, perché risponde all'esigenza, da ultimo ben evidenziata dal Parlamento europeo, di tracciare le linee politiche fondamentali, anche in termini di divieti generalizzati, all'interno delle quali potrà trovare espressione l'utilizzo dei nuovi strumenti tecnici. Quello

dell'*information technology*, infatti, è un universo la cui capacità di evoluzione è talmente rapida, da rendere qualsiasi disciplina normativa di dettaglio, anche la più attenta ed aggiornata, fisiologicamente arretrata rispetto allo sviluppo della tecnica e in quanto tale inidonea ad assicurare strumenti effettivi di governo del fenomeno.