

TRA ALGORITMO E TOGA: EX-AEQUO E COMPAS COME LABORATORI DI CONVIVENZA TRA INTELLIGENZA ARTIFICIALE E GIUSTIZIA PENALE

Giuseppe Tuccillo*

SOMMARIO: 1.- Introduzione; 2.- “Compas” ed “Ex-Aequo”: l’ingresso dell’algoritmo nella valutazione penale; 3.- Giustizia algoritmica e dosimetria penale: un confronto tra i modelli Compas ed Ex-Aequo; 4.- Il problema dell’accertamento del nesso causale con le implicazioni algoritmiche; 5.- Governare l’intelligenza artificiale nella giurisdizione penale: prospettive *de iure condendo*; 6.- Giustizia e fiducia: verso una convivenza costituzionalmente sostenibile tra diritto e intelligenza artificiale.

1.- Introduzione.

L’adozione delle tecnologie basate sui sistemi di Intelligenza Artificiale¹ segna, senza dubbio, una delle più profonde e pervasive metamorfosi nella storia della società contemporanea, destinata a incidere in modo strutturale su settori nevralgici dell’ordinamento, tra cui, in posizione non marginale, quello giudiziario.

Negli ultimi anni, infatti, si è assistito ad una progressiva evoluzione dei sistemi intelligenti, il cui impiego, nel contesto giuridico, appare ormai suscettibile di ridisegnare — in maniera non episodica — l’intero “ecosistema” della giurisdizione, dalla gestione automatizzata della documentazione processuale alla consulenza legale predittiva, sino a lambire le fondamenta del processo decisionale. Se da un lato l’introduzione di strumenti algoritmici avanzati, segnatamente quelli fondati su “machine learning” (ML) e “natural language processing” (NLP), parrebbe idoneo a promuovere un incremento dell’efficienza dell’attività giudiziaria, dall’altro tale impiego solleva una pluralità di problematiche etiche e sociali, tutt’altro che recessive².

L’impiego di meccanismi di calcolo predittivo e di analisi automatizzata del linguaggio giuridico non può prescindere, infatti, da una riflessione approfondita sui rischi derivanti dall’opacità algoritmica, dall’eventuale compromissione del principio di non discriminazione, dalla difficoltà di imputazione della responsabilità in caso di esiti erronei, nonché dalle implicazioni derivanti dal trattamento massivo di dati personali, anche di natura sensibile.

In tale quadro, si impone un’indagine rigorosa, connessa all’introduzione dell’IA in ambito giuridico, nella consapevolezza che l’innovazione tecnologica, per poter essere accolta in modo legittimo, debba conformarsi ai principi fondamentali dell’ordinamento giuridico ed alle garanzie costituzionalmente tutelate³.

* Dottorando di ricerca in Diritto Penale presso l’Università degli Studi di Salerno.

¹ D’ora in poi anche “I.A.”.

² R.E. Kistoris, *Intelligenza artificiale, strumenti predittivi e processo penale*, in *Cass. Pen.*, 2024, p. 1647 ss.

³ Per una più approfondita analisi sulle “luci” e le “ombre” dei sistemi di I.A. in ambito giuridico, si leggano- *ex multis*- C. Barbaro, *Uso dell’intelligenza artificiale nei sistemi giudiziari: verso la definizione di principi etici condivisi a livello europeo?*, in *Quest. Giust.*, 4/2018, 195; E. Bassoli, *Algoritmica giuridica. Intelligenza artificiale e diritto*, Amon, 2021; G. Carcaterra, *Presupposti e strumenti della scienza giuridica*, Giappichelli, 2011; F. Consulich, M. Maugeri, T.N. Poli, G. Trovatore, *AI e abusi di mercato: le leggi della robotica si applicano alle operazioni finanziarie?*, in *Quad. CONSOB*, 29 maggio 2023, 29 ss.; A. D’Aloia, *Intelligenza artificiale e diritto. Come regolare un mondo nuovo*, Franco Angeli, 2021; M.R. Magliulo, *L’Intelligenza Artificiale nel processo penale: progresso o rischio per la tutela dei diritti costituzionali?* in *Il processo*, 2022, 2, 559 ss; C. Parodi, V. Sellaroli, *Sistema penale e intelligenza artificiale: molte speranze e qualche equivoco*, in *Diritto Penale Contemporaneo*, Fascicolo 6/2019, pp. 47-71; S. Riondato, *Robot: talune*

Il presente contributo si prefigge, dunque, l'obiettivo di esplorare le principali applicazioni dell'IA in ambito giuridico, con particolare riguardo ai sistemi "Compas" ed "Ex-Aequo" quali ausili alla funzione giudiziaria.

2.- "Compas" ed "Ex-Aequo": l'ingresso dell'algoritmo nella valutazione penale.

L'adozione di sistemi algoritmici a supporto della funzione giurisdizionale promette un incremento dell'efficienza operativa, ma implica un radicale ripensamento dei presupposti del giudizio penale, in particolare nei momenti ad alta discrezionalità, come la valutazione del rischio di recidiva ovvero la commisurazione della pena.

Tra gli strumenti più noti in tal senso si colloca il modello COMPAS⁴, progettato per elaborare previsioni basate su dati socio-criminologici e comportamentali. Il sistema, strutturato su parametri non interamente accessibili, opaco sotto il profilo procedurale, è stato oggetto di numerose critiche in merito alla sua capacità di garantire equità predittiva, trasparenza e rispetto delle garanzie fondamentali dell'imputato⁵. L'uso di dati potenzialmente errati e la difficoltà di controllare, in modo indipendente, le scelte dell'algoritmo rischiano di trasformarlo da semplice supporto a elemento che condiziona la decisione del giudice⁶.

È nel delicato passaggio tra valutazione prognostica - sulla quale si tornerà più approfonditamente nei paragrafi successivi - e commisurazione della sanzione che l'interazione tra IA e giurisdizione richiede un approfondimento critico, volto a distinguere modelli eterogenei per impostazione e tenuta garantista⁷.

Nel contesto delle sperimentazioni algoritmiche applicate alla giurisdizione penale, il modello Ex-Aequo rappresenta un prototipo di supporto decisionale concepito in aderenza ai principi costituzionali e convenzionali che presidiano la funzione giudicante. Il sistema si propone di affiancare il giudice nella fase di commisurazione della pena, mediante un impianto metodologico strutturato in tre fasi: raccolta e standardizzazione dei dati giurisprudenziali, elaborazione algoritmica fondata su analogie fattuali, e restituzione di un output informativo⁸.

L'architettura di Ex-Aequo è pensata per offrire un supporto alla comprensione e alla valutazione del giudice in cui l'algoritmo non opera in funzione predittiva né sostitutiva, ma come strumento di raffronto e affinamento del giudizio umano. La sua funzione è rendere accessibile, in forma trasparente e motivata, l'andamento della prassi sanzionatoria consolidata, restituendo al giudice un intervallo statistico delle pene irrogate in casi analoghi, insieme agli eventuali elementi di discontinuità. Elemento qualificante del modello è la riserva di decisione umana, garantita sia sul

implicazioni di diritto penale, in *Tecnodiritto. Temi e problemi di informatica e robotica giuridica*, a cura di Moro-Sarra, Milano, 2017, 85ss.

⁴ Il cui nome, per esteso, è "Correctional Offender Management Profiling for Alternative Sanctions".

⁵ In argomento, *ex multis*, O. Di Giovine, *Il judge-bot e le sequenze giuridiche in materia penale (intelligenza artificiale e stabilizzazione giurisprudenziale)*, in *Giur. pen.*, 2020, 3, 955; L. Maldonato, *Algoritmi predittivi e discrezionalità del giudice: una nuova sfida per la giustizia penale*, in *Dir. pen. cont.*, 2019, 2, p. 404 ss.; 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.*, 2021. C. Parodi-V. Sellaroli, *op. cit.*, 2019, 6, p. 54.

⁶ G. Sartor, F. Lagioia, *Il sistema COMPAS: algoritmi, previsioni, iniquità*. Torino, Giappichelli 2021.

⁷ V. Manes, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in *Discrimen*, 15 maggio 2020; J. Nieva-Fenoll, *Intelligenza artificiale e processo*, Torino, Giappichelli, 2019, p. 109; S. Quatroccolo, *Equo processo penale e sfide della società algoritmica*, in *BioLaw Journal – Rivista di BioDiritto*, n. 1, 2019, pp. 135-144.

⁸ Si veda F. Coppola, *Commisurazione della pena e intelligenza artificiale: una ipotesi di lavoro con l'algoritmo Ex-Aequo*, in *Arch. Pen.*, 2/2023.

piano formale che funzionale: l'output algoritmico costituisce uno stimolo cognitivo ma non vincola né condiziona l'esito sanzionatorio⁹.

Tale impostazione intende prevenire fenomeni di automatismo valutativo, incentivare l'esplicitazione dei criteri di commisurazione e favorire un uso riflessivo della tecnologia in ambito giurisdizionale¹⁰. A fronte di tali criticità, si rende opportuno un confronto tra modelli algoritmici alternativi, basati su presupposti epistemologici e meccanismi di garanzia differenti, come COMPAS ed Ex-Aequo. È a questo confronto che si dedica il paragrafo che segue.

3.- Giustizia algoritmica e dosimetria penale: un confronto tra i modelli Compas ed Ex-Aequo.

L'impiego dell'intelligenza artificiale nelle decisioni della giustizia penale solleva interrogativi centrali, sia in termini di legittimità sia di opportunità. In tale contesto, si impone un'analisi comparativa tra due modelli emblematici, già descritti nel paragrafo precedente: il sistema Compas, utilizzato negli Stati Uniti, e il più recente progetto Ex-Aequo.

Pur condividendo l'obiettivo di automatizzare parte dell'elaborazione cognitiva connessa alla commisurazione della pena, i due sistemi si distinguono per impostazione metodologica e garanzie a tutela dei diritti fondamentali.

Il "caso Compas", reso celebre dalla sentenza della Corte Suprema del Wisconsin nel 2016¹¹, costituisce uno snodo paradigmatico. In quella sede, l'algoritmo venne impiegato nella determinazione della pena, sulla base della previsione di rischio di recidiva fondata su variabili soggettive e socio-demografiche. Il giudice, recependo il risultato generato dal sistema – che assegnava all'imputato un rischio elevato – comminava una pena significativamente più severa¹².

Sebbene la Corte Suprema abbia confermato la legittimità dell'utilizzo dello strumento, ha evidenziato come i risultati predittivi non possano costituire l'unico fondamento della decisione giudiziale. Essa ha sottolineato la necessità di un impiego "cautelato" dell'algoritmo, escludendo ogni automatismo e affermando il primato della valutazione del giudice.

Nondimeno, la vicenda ha portato alla luce criticità quali l'opacità dei criteri interni dell'algoritmo e la difficoltà di verificare *ex post* il processo logico seguito.

Altro problema è rappresentato dal rischio di riproduzione o amplificazione di bias discriminatori, etnico-razziali.

In netta controtendenza rispetto a tale modello si colloca Ex-Aequo.

Tale sistema nasce esplicitamente nel solco del principio secondo cui la decisione giudiziale deve rimanere appannaggio dell'intelligenza umana, dotata di sensibilità valutativa, responsabilità istituzionale e capacità di contestualizzazione.

L'elemento distintivo risiede nella scelta metodologica di anteporre l'elaborazione della ricerca algoritmica alla decisione del giudice, così da prevenire ogni uso strumentale del sistema e ogni rischio di "manipolazione a posteriori"¹³.

⁹ Per un'accurata e puntuale analisi sul "sistema Ex-Aequo", si veda F. Coppola, *Intelligenza Artificiale, Metaverso e Sistema Penale: prevenzione, repressione, opportunità, rischi*, Cedam, 2025, pp. 153-164.

¹⁰ A. Bondi, G. Marra – R. Palavera (a cura di), *Diritto penale tra teoria e prassi*, Urbino University Press, 2024, 119 ss.

¹¹ Cfr. *Wisconsin Supreme Court, State v. Loomis*, case 2015AP157-CR, Judgement July 13th 2016, in *Harvard LR*, 2017, 130, 1530 ss.

¹² A. Calcaterra, *Machinae autonome e decisione robotica*, in A. Carleo (a cura di), *Decisione robotica*, Il Mulino, Bologna, 2019; S. Carrer, *Se l'amicus curiae è un algoritmo: il chiacchierato caso Loomis alla Corte Suprema del Wisconsin*, in *Giurisprudenza Penale Web*, 2019, 4.

¹³ Coppola, *op.cit.*, pp. 156-164.

A differenza di Compas, l'obiettivo di Ex-Aequo è quello di restituire al giudice un "range" di pene statisticamente prevalenti in casi analoghi, accompagnato da un'analisi delle variabili rilevanti e delle eventuali circostanze eccentriche che giustificano una divergenza.

L'algoritmo funge da strumento di ausilio alla motivazione, ma non vincola l'esito.

Non si può tuttavia ignorare che anche Ex-Aequo, come ogni sistema algoritmico, non è immune da rischi¹⁴.

Il rischio che l'algoritmo assorba pregiudizi presenti nelle decisioni giudiziarie passate, anche se formalmente legittime, rende necessario un monitoraggio costante. Per limitarlo, il modello prevede due garanzie: da un lato, il giudice mantiene "l'ultima parola"¹⁵, potendo correggere errori e considerare nuovi elementi; dall'altro, l'algoritmo è in grado di riconoscere decisioni isolate o distorte, trattandole come eccezioni, e valorizzare quelle più ricorrenti come segnali di coerenza.

Infine, una differenza sostanziale rispetto al modello Compas riguarda il rammentato principio di trasparenza e verificabilità.

Orbene, il confronto tra COMPAS ed Ex-Aequo evidenzia due approcci opposti alla giustizia algoritmica.

Il primo, tipico del contesto anglosassone, assegna all'algoritmo compiti decisionali ad alto contenuto discrezionale¹⁶.

Il secondo è ispirato- *mutatis mutandis*- ai principi europei e costituzionali e vede l'intelligenza artificiale come uno strumento di supporto, sempre sotto il controllo del giudice, pensato per intestare coerenza alla decisione giudiziale, senza rinunciare però alla componente umana della giurisdizione.

4.- Il problema dell'accertamento del nesso causale con le implicazioni algoritmiche.

La riflessione sui modelli di giustizia algoritmica impone di interrogarsi su come l'intelligenza artificiale incida sulla ricostruzione del nesso causale, presupposto imprescindibile per l'attribuzione della responsabilità penale. Nel diritto penale, la causalità è funzionale a circoscrivere la responsabilità individuale, costituendo un elemento strutturale del reato, soprattutto nelle fattispecie di evento. Tradizionalmente, la dottrina ha elaborato diversi modelli teorici per affrontare il problema del nesso eziologico: la teoria condizionalistica (o della *conditio sine qua non*), basata sul criterio controfattuale dell'eliminazione mentale dell'antecedente; la teoria della causalità adeguata, che seleziona le cause penalmente rilevanti in base alla regolarità causale; la teoria dell'imputazione oggettiva, che valuta la creazione di un rischio giuridicamente rilevante e la sua concretizzazione nell'evento¹⁷.

Nonostante le differenze, tutte queste ricostruzioni condividono un presupposto implicito: la riferibilità del processo causale a un soggetto umano capace di prevedere e controllare gli effetti della

¹⁴ Coppola, *Commisurazione della pena e intelligenza artificiale: una ipotesi di lavoro con l'algoritmo Ex-Aequo*, cit.

¹⁵ B. Occhiuzzi, "Algoritmi predittivi: alcune premesse metodologiche", in *Diritto penale contemporaneo – Rivista trimestrale*, n. 2, 2019, pp. 391-400; *Giocare con altri dadi Giustizia e predittività dell'algoritmo*, Mastroicavo (a cura di), Giappichelli, 2024, p. 138-143.

¹⁶ Per una più specifica analisi sulle implicazioni dell'utilizzo dell'intelligenza artificiale nel processo penale, si veda *Intelligenza artificiale e processo penale. Indagini, prove, giudizio*, a cura di Di Paolo-Pressacco, Trento, 2022; Donati-Fiorelli, *Diritto probatorio e giudizi criminali ai tempi dell'intelligenza artificiale*, in *Dir. pen. cont.*, 2022, 2, 24 ss.; M. Pisati, "Indagini preliminari e intelligenza artificiale: efficienza e rischi per i diritti fondamentali", in *Processo penale e giustizia*, 2020, pp. 957-971; G. Riccio, "Ragionando su intelligenza artificiale e processo penale", in *Archivio penale – Rivista web*, n. 3, pp. 1-13; L. Romanò, *Intelligenza artificiale come prova scientifica nel processo penale*, in G. Canzio – L. Lupária (a cura di), *Prova scientifica e processo penale*, CEDAM, Padova, 2022.

¹⁷ Cfr. *ex multis*, G. Marinucci – E. Dolcini – G. L. Gatta, *Manuale di diritto penale. Parte Generale*, IX ed., Giuffrè, 2020, p. 231-243; F. Antolisei, *Manuale di diritto penale. Parte Generale*, XVI ed., Giuffrè, 2003, pp. 198-198-216.

propria condotta. È infatti causalmente rilevante, in senso penalistico, solo quell'antecedente che ha generato un rischio tipico concretizzatosi nell'evento.

Tuttavia, l'emergere di sistemi decisionali automatizzati e l'utilizzo crescente di tecniche di apprendimento statistico impongono una revisione concettuale del nesso causale. L'autonomia operativa e l'opacità algoritmica, invero, minano la linearità del rapporto causa-effetto, rendendo difficile individuare un unico soggetto responsabile. L'evento lesivo può derivare non da una condotta umana immediata, ma da scelte progettuali compiute a monte o da carenze di addestramento che generano risultati imprevedibili. In questi casi, il nesso si frammenta e si distribuisce lungo una catena causale complessa, che coinvolge sviluppatori, data scientist, fornitori e utilizzatori finali¹⁸.

Il tradizionale test controfattuale ("se A non fosse accaduto, B si sarebbe verificato?") perde di efficacia, perché i processi decisionali dell'intelligenza artificiale non sono sempre replicabili né pienamente comprensibili, neppure dagli stessi progettisti.

L'intervento dell'intelligenza artificiale compromette la linearità del nesso eziologico nella parte in cui la catena causale si spezza e si frammenta, rendendo difficile individuare un singolo autore o una condotta direttamente imputabile.

L'incertezza sul "che cosa" abbia realmente causato l'evento, e se esso sia riferibile a una condotta umana, rende problematica l'imputazione soggettiva della responsabilità¹⁹.

Inoltre, la difficoltà di accertamento si aggrava se si considera che l'attività giudiziale, specie nelle valutazioni eziologiche *ex post* e prognostiche *ex ante*, si fonda su risorse cognitive limitate e su processi decisionali esposti a distorsioni²⁰.

Invero, la diagnosi consiste nell'accertamento *ex post* delle cause di un determinato evento e si fonda sull'analisi delle evidenze disponibili e delle leggi scientifiche note. La prognosi, invece, attiene alla previsione di comportamenti futuri e si esprime in termini probabilistici, come accade nella valutazione della pericolosità sociale, della recidiva o ancora nella concessione di benefici quali la sospensione condizionale della pena²¹.

I bias cognitivi possono compromettere l'oggettività del giudizio, influenzando la ricostruzione della catena causale. Tra questi, la fallacia della regola cd. "di inusualità", che considera falsa un'ipotesi solo perché l'evento è ritenuto troppo raro per essere casuale: tal ricostruzione può indurre in errore. Il caso emblematico è quello di Sally Clark, condannata ingiustamente per l'omicidio dei figli sulla base di una valutazione statistica errata, dimostra come tali trappole cognitive possano compromettere gravemente il giudizio penale²².

¹⁸ F. Basile, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Diritto Penale e Uomo*, 10/2019, 1, 10; Carcaterra, *op.cit.*, pp. 245 ss.; G. Forti, *Colpa ed evento nel diritto penale*, Milano, 1990; M.B. Magro, *Decisione umana e decisione robotica. Un'ipotesi di responsabilità da procreazione robotica*, in *Leg. pen.*, 10 maggio 2020, p. 12; A. Massaro, *Principio di precauzione e diritto penale: nihil novi sub sole?*, in *Dir. pen. cont.*, 2011; C. Piergallini, *Intelligenza artificiale: da "mezzo" ad "autore" del reato?* in *Riv. it. dir. proc. pen.*, n.4, 2020, 1745 ss.

¹⁹ A. Gargani, *La "flessibilizzazione" giurisprudenziale delle categorie classiche del reato di fronte alle esigenze di controllo penale delle nuove fenomenologie di rischio*, in *Leg. pen.*, 2011, n. 2, p. 397 ss.; B. Fragrasso, *La responsabilità penale del produttore di sistemi di intelligenza artificiale*, in *Sistema penale*, 2023, pp. 19-22; B. Panattoni, *Intelligenza artificiale: le sfide per il diritto penale nel passaggio dall'automazione tecnologica all'autonomia artificiale*, in *Dir. inf.*, 2021, fasc. 1; L. Picotti, *Profili di responsabilità penale per la circolazione di veicoli a guida autonoma*, in M. Catenacci – V. D'Ascola – R. Rampioni (a cura di), *Studi in Onore di Antonio Fiorella*, I, RomaTre Press, 2021, 827-828.

²⁰ D. Perrone, *La prognosi postuma tra distorsioni cognitive e software predittivi. Limiti e possibilità del ricorso alla "giustizia digitale integrata" in sede di accertamento della colpa*, Giappichelli, Torino 2021, pp. 43-59.

²¹ G. De Francesco, *Colpa e prevenzione del rischio nel campo delle malattie professionali*, in *Dir. pen. proc.*, 2012, 665 ss.; Forti, *op.cit.*, Milano, 1990, 200 ss.; C. Perini, *Il concetto di rischio nel diritto penale moderno*, Milano, 2010, 367 ss.

²² In argomento, per una più approfondita analisi, Perrone, *op.cit.*, pp.37-45.

Questi meccanismi, spesso inconsci, incidono sulla ricostruzione delle effettive capacità rappresentative e previsionali dell'agente al momento del fatto, portando a sovrastimare la possibilità di prevedere o impedire l'evento. Si rischia così di fondare l'imputazione su un nesso causale oggettivo, senza un adeguato coefficiente soggettivo di rimproverabilità. Tale slittamento può compromettere la valutazione della colpevolezza, spingendo il giudice a ritenere esigibile una condotta diversa da parte dell'agente sulla base di aspettative cognitive distorte.

Ciò solleva interrogativi cruciali: può ancora reggere un modello di responsabilità penale fondato su dolo o colpa, quando la catena causale si radica in sistemi autonomi e sfugge al controllo individuale? La difficoltà di inquadrare l'azione dell'intelligenza artificiale negli schemi classici della causalità penalistica impone una profonda riconsiderazione del ruolo del diritto penale nell'epoca digitale²³.

Occorre sviluppare strumenti concettuali capaci di fare i conti con l'incertezza epistemologica, di prevenire l'oggettivazione dei bias cognitivi nel giudizio e di ristabilire un equilibrio tra razionalità predittiva e garanzie soggettive.

5.- Governare l'intelligenza artificiale nella giurisdizione penale: prospettive *de iure condendo*.

L'introduzione progressiva di strumenti algoritmici nei processi decisionali della giustizia penale non costituisce un mero avanzamento tecnico, bensì una trasformazione paradigmatica, idonea a incidere sulle basi stesse del sistema penale di matrice liberal-garantista.

La funzione giurisdizionale, infatti, non può ridursi ad una "meccanica applicazione di norme ai fatti", ma deve essere intesa come attività di giudizio, intrinsecamente legata alla discrezionalità del decisore e alla capacità di argomentare pubblicamente il proprio convincimento²⁴.

In tale contesto, si impone la necessità di un intervento normativo *de iure condendo*, volto a delineare un quadro giuridico chiaro per la regolazione dell'intelligenza artificiale in ambito penale. Occorre, invero, anzitutto un intervento "netto", ad opera del legislatore, prodromico a sancire in modo esplicito il carattere consultivo e non vincolante degli strumenti algoritmici, affermando che nessun algoritmo può sostituirsi al giudizio umano nella valutazione del fatto o nella determinazione della pena²⁵.

Inoltre, è essenziale definire standard stringenti di trasparenza ed accessibilità: le scelte progettuali e metodologiche devono essere documentate; le fonti dei dati devono poter essere verificate.

In assenza di queste garanzie, si corre il rischio di introdurre nel processo penale una "black box epistemica", radicalmente incompatibile con il principio di legalità sostanziale e con il diritto dell'imputato a conoscere le ragioni del trattamento sanzionatorio cui è sottoposto²⁶.

²³ Di Giovine, *op. cit.*, 959; Massaro, *Intelligenza artificiale e neuroscienze: l'eterno ritorno del diritto penale?*, in J.M. Peris Riera-Massaro (a cura di), *Diritto penale, intelligenza artificiale e neuroscienze*, RomaTre Press, 2023, p. 46.

²⁴ Cfr., *ex multis*, Basile, *op.cit.*, p.8; Coppola, *op.cit.*, pp. 1-28; Manes, *op.cit.*; R. Borsari, *Intelligenza Artificiale e responsabilità penale: prime considerazioni*, in *Medialaws*, 2020; Di Giovine, *op.cit.*; M. Papa, *Future crimes: intelligenza artificiale e rinnovamento del diritto penale*, in *Criminalia*, 2019.

²⁵ Sulla cd. "tollerabilità sociale" dei rischi prospettabili con l'avvento delle I.A. generative che si accompagneranno a decisioni "politiche" cfr. Coppola, *op. cit.* p. 169-170. Al riguardo, l'autore ritiene che l'approccio più verosimile sarà "gattopardista: questa opzione, che, prescindendo per un momento dalle possibili modalità di declinazione, ci appare quella più verosimile, potrebbe mettere 'a frutto' entrambe le precedenti impostazioni. Da un lato, strizzare l'occhio al progresso, consentendo la sperimentazione dei sistemi di I.A. progrediti; dall'altro, assicurare l'elettorato con fattispecie avamposto contro i pericoli generati dalla macchina."

²⁶ Per una più accurata analisi sui rischi derivanti da I.A. generativa, G. Balbi – F. De Simone Andreana Esposito – S. Manacorda (a cura di), *Diritto penale e intelligenza artificiale, "Nuovi scenari"*, G. Giappichelli Editore – Torino 2022 p. 68; C. Brusco, *Rischio e pericolo, rischio consentito e principio di precauzione. La c.d. "flessibilizzazione delle categorie del reato"*, in *Criminalia* 2012, 383- 414; Consulich, voce *Rischio consentito*, in *Enciclopedia del diritto. Reato colposo*, diretto da M. Donini, Milano 2021, 1102 ss; Gargani, *op.cit.*, 397 ss.; F. Herzog, *Società del rischio, diritto*

È altresì necessario promuovere una formazione multidisciplinare dei magistrati, affinché possano impiegare consapevolmente gli strumenti predittivi, valutandone criticamente potenzialità e limiti²⁷. La conoscenza tecnica non rappresenta un elemento accessorio, ma una condizione essenziale per l'esercizio pieno e responsabile della giurisdizione nell'era digitale. In questa prospettiva, la cultura giuridica deve partecipare attivamente alla progettazione delle architetture algoritmiche, per evitare che la giurisdizione venga guidata da modelli tecnici opachi e sottratti al controllo democratico. Infine, qualsiasi integrazione dell'IA nel processo penale deve muoversi entro una cornice costituzionale salda: il rispetto dei principi di personalità della pena, ragionevolezza, eguaglianza, proporzionalità e colpevolezza non può essere sacrificato sull'altare dell'efficienza. Solo in tal modo essa potrà risultare compatibile con un modello di giustizia penale fedele alla propria ispirazione garantista e ai fondamenti costituzionali che la sorreggono²⁸.

6.- Giustizia e fiducia: verso una convivenza costituzionalmente sostenibile tra diritto e intelligenza artificiale.

Il diritto, per sua natura, cerca di dare forma giuridica ai rapporti tra le persone, trasformando la complessità delle azioni individuali e collettive in regole socialmente rilevanti. In questo contesto, l'intelligenza artificiale non è solo un progresso tecnico, ma può modificare in profondità il modo in cui il diritto si concepisce, incidendo sia sul funzionamento delle pratiche giuridiche sia sul rapporto tra individuo e ordinamento.

L'introduzione di sistemi algoritmici nella giurisdizione penale solleva, dunque, un interrogativo fondamentale: che tipo di giustizia desideriamo costruire in una società digitale²⁹?

Non è sufficiente chiedersi “come” impiegare la tecnologia, ma occorre interrogarsi sul “perché” essa debba assumere un ruolo così pervasivo in un ambito delicato come quello del giudizio penale. In tale prospettiva, la questione dell'intelligenza artificiale non può essere ridotta a un problema di efficienza amministrativa o di razionalizzazione delle risorse, ma deve essere affrontata come una questione di legittimazione democratica e di fiducia sociale nel diritto.

Il principio di fiducia, peraltro, non è un elemento secondario del sistema giuridico, ma rappresenta una componente strutturale del “patto di convivenza democratica”: il cittadino riconosce il potere punitivo dello Stato nella misura in cui confida che esso sia esercitato da esseri umani razionali, imparziali, responsabili, capaci di interpretare le norme alla luce dei valori costituzionali e delle mutevoli esigenze sociali. Tale fiducia è radicata in una visione antropocentrica della giustizia, dove il giudice è chiamato a comprendere, oltre che valutare, a sentire, oltre che argomentare, operando una sintesi di ragione e coscienza³⁰.

L'algoritmo, per quanto sofisticato, resta un'entità priva di empatia e comprensione morale: esso non conosce la colpa ed il pentimento.

penale del rischio, regolazione del rischio, in Critica e giustificazione del diritto penale nel cambio di secolo. L'analisi critica della Scuola di Francoforte, a cura di L. Stortoni-L. Foffani, Milano 2004, 357; Perini, *op.cit.*, Milano 2010.

²⁷ Coppola, *op.cit.*, p. 157.

²⁸ Cfr. A. Scalfati, *IA e processo penale: prospettive d'impiego e livelli di rischio*, in *Processo penale e giustizia*, n. 6/2024.

²⁹ P. Moro, “Macchine come noi. Natura e limiti della soggettività robotica”, in Ruffolo, Ugo (editor): *Intelligenza artificiale. Il diritto, i diritti, l'etica*, Milano, Giuffrè Francis Lefebvre, 2020, pp. 45-61; G. Padua, *Intelligenza artificiale e giudizio penale: scenari, limiti e prospettive*, in *Processo penale e giustizia*, n. 6/2021; G. Uberris, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, in *Diritto penale contemporaneo – Rivista trimestrale*, n. 4/2020.

³⁰ Manes, *Introduzione ai principi costituzionali in materia penale*, Giappichelli, 2023, pp. 36-38.

Anche quando opera correttamente sul piano statistico, l'algoritmo non conosce il caso concreto: conosce una media o una probabilità. È quindi ontologicamente inidoneo a sostituirsi alla funzione giurisdizionale, che è per definizione assiologicamente orientata.

Ciò non significa però rifiutare l'IA in modo pregiudiziale. Al contrario, essa può offrire un valido supporto verso un crescente sviluppo, un "consulente invisibile" capace di ampliare l'orizzonte della conoscenza del giudice, fornendo soluzioni senza scavalcare la valutazione discrezionale. Il vero rischio, infatti, non risiede nell'impiego dell'algoritmo in sé, ma nell'opacità dei criteri che lo governano e nella possibilità che il decisore umano abdichi alla propria responsabilità, trincerandosi dietro l'autorità del calcolo³¹.

È per questa ragione che ogni futura regolamentazione dell'IA nel settore penale dovrà farsi carico della protezione del principio di soggettività della decisione giurisdizionale, predisponendo garanzie effettive di trasparenza e controllo sull'uso degli strumenti tecnologici.

Non si tratta solo di stabilire dei limiti esterni all'IA, ma di costruire un modello di "co-decisione responsabile", dove l'apporto algoritmico sia integrato nel ragionamento giuridico senza soffocarlo. In definitiva, il progresso tecnologico offre l'opportunità di rinnovare la funzione giudiziaria, senza snaturarla. Ma perché ciò avvenga, è necessario che il diritto non smarrisca la propria anima antropologica.

La giustizia penale, infatti, non è misurabile in bit o in millisecondi. Solo in questo equilibrio dinamico sarà possibile cogliere al meglio le opportunità offerte dall'IA, di cui Ex-Aequo è fulgido esempio, verso una giurisdizione all'altezza delle sfide future.

Abstract.- Il contributo affronta la progressiva integrazione dell'intelligenza artificiale nella giurisdizione penale, con particolare riguardo al confronto tra due modelli paradigmatici: COMPAS, utilizzato negli Stati Uniti, ed Ex-Aequo, sviluppato in ambito europeo. Pur accomunati dall'intento di assistere il giudice nella valutazione sanzionatoria, i sistemi divergono radicalmente: il primo opera in chiave predittiva, opaca e potenzialmente discriminatoria; il secondo si fonda su trasparenza metodologica, restituisce intervalli sanzionatori desunti da casi analoghi e mantiene intatta la centralità del libero convincimento giudiziale. L'analisi si estende al problema del nesso causale, che l'opacità algoritmica tende a frammentare, sollevando interrogativi sulla tenuta dei criteri classici di imputazione penale e sul rischio di oggettivazione dei bias cognitivi. In prospettiva de iure condendo, si sottolinea l'urgenza di una cornice normativa che sancisca il ruolo consultivo degli algoritmi, definisca standard stringenti di trasparenza e promuova la formazione tecnico-giuridica dei magistrati. In conclusione, si prospetta un modello di "co-decisione responsabile", in cui l'IA rafforza la motivazione senza sostituire la sensibilità valutativa del giudice, preservando così fiducia e legittimazione democratica della giustizia penale.

The paper addresses the progressive integration of artificial intelligence into criminal jurisdiction, with particular regard to the comparison between two paradigmatic models: COMPAS, used in the United States, and Ex-Aequo, developed within the European context. While both share the aim of assisting the judge in sentencing assessment, the systems diverge radically: the former operates in a predictive, opaque, and potentially discriminatory manner; the latter is grounded in methodological

³¹ U. Ruffolo, *Intelligenza Artificiale, machine learning e responsabilità da algoritmo*, in *Giurisprudenza italiana*, 2019, pp. 1689-1704; A. Simoncini, *Diritto costituzionale e decisioni algoritmiche*, in Dorigo, Stefano (editor): *Il ragionamento giuridico nell'era dell'intelligenza artificiale* (Pisa, Pacini), 2020, pp. 37-65; E. Vincenti, "Il «problema» del giudice-robot", in Carleo, Alessandra (editor): *Decisione robotica* (Bologna, il Mulino), 2019, pp. 111-124.

transparency, provides sanctioning ranges inferred from analogous cases, and preserves the centrality of the judge's free conviction. The analysis extends to the issue of causality, which algorithmic opacity tends to fragment, raising questions about the resilience of classical criteria of criminal imputation and the risk of objectifying cognitive biases. From a *de iure condendo* perspective, the urgency emerges of a regulatory framework that enshrines the consultative role of algorithms, establishes stringent transparency standards, and promotes the technical-legal training of judges. In conclusion, the paper envisages a model of "responsible co-decision," in which AI strengthens judicial reasoning without replacing the judge's evaluative sensitivity, thereby preserving trust and the democratic legitimacy of criminal justice.